

REVISTA

DA UNIVERSIDADE FEDERAL
DE MINAS GERAIS



#30

ISSN 2316-770X

A Revista da Universidade Federal de Minas Gerais é uma publicação quadrimestral e tem o objetivo principal de abordar temáticas específicas, numa perspectiva interdisciplinar, podendo divulgar também resultados de pesquisas e de produções teóricas e artísticas diversas.

GESTÃO 2022-2026

Sandra Regina Goulart Almeida
REITORA

Alessandro Fernandes Moreira
VICE-REITOR

Rui Rothe-Neves
CHEFE DE GABINETE

Ivan José da Silva Lopes
PRÓ-REITOR DE ADMINISTRAÇÃO

Claudia Andréa Mayorga Borges
PRÓ-REITORA DE EXTENSÃO

Bruno Otávio Soares Teixeira
PRÓ-REITOR DE GRADUAÇÃO

Fernando Marcos dos Reis
PRÓ-REITOR DE PESQUISA

Maurício Freire Garcia
PRÓ-REITOR DE PLANEJAMENTO E
DESENVOLVIMENTO

Isabela Almeida Pordeus
PRÓ-REITORA DE PÓS-GRADUAÇÃO

Maria Márcia Magela Machado
PRÓ-REITORA DE RECURSOS HUMANOS

Tarcísio Mauro Vago
PRÓ-REITOR DE ASSUNTOS ESTUDANTIS

Estevam Barbosa de Las Casas
DIRETOR DO IEAT - UFMG

Revista da UFMG

EDITORA-CHEFE
Ana Maria Rabelo Gomes

EDITORES CONVIDADOS PARA ESTA EDIÇÃO

Frederico Gadelha Guimarães
Estevam Barbosa de Las Casas
André Fujita
Toshio Fukuda

CURADORIA ARTÍSTICA
Isabela Sales Prado

EDITOR EXECUTIVO
Lorena Tavares de Paula

SECRETARIA EXECUTIVA
Tatiana Pereira Queiroz

SECRETARIA EDITORIAL
Daniele Catizane

PORTAL DE PERIÓDICOS UFMG
Carla Oliveira

FICHA CATALOGráfICA

R 454 Revista da Universidade Federal de Minas Gerais. –
vol.15, 1965- – Belo Horizonte : UFMG, 1965- v. : il.
Quadrimestral, 2021-
Anual de 1965-1969.
A partir do v. 27, n. 1, 2021, passa a ser quadrimestral.
A partir do v. 19, n. 1/2, 2012 passa a ser semestral.
Título anterior: Revista da Universidade de Minas Gerais, 1929-
1964.
ISSN 2316-770X.
Disponível em: <<https://periodicos.ufmg.br/index.php/revista-daufmg/index>>.
Quadrimestral
Inclui bibliografia.
I. Ensino superior- Periódicos. I. Universidade Federal de Minas Gerais.

CDD: 378.405 CDU: 378

Elaborada pela DITI – Setor de Tratamento da Informação
Biblioteca Universitária da UFMG.
Bibliotecário Rafael Gonçalves Dias. CRB: MG-003277/O.

PROJETO GRÁFICO

Leo Ruas

PRODUÇÃO GRÁFICA

Warren Marilac, Editora da UFMG

DIAGRAMAÇÃO

Lucas Matoso, Estagiário do IEAT

REVISÃO

Agência Rivera de Consultoria em Comunicação

PRODUÇÃO EXECUTIVA

Instituto de Estudos Avançados Transdisciplinares da UFMG

Conselho editorial

Alfredo González-Ruibal •
INSTITUTO DE CIÊNCIAS DO PATRIMÔNIO, ESPANHA

Anísio Brasileiro de Freitas Dourado •
UNIVERSIDADE FEDERAL DE PERNAMBUCO - UFPE, BRASIL.

Edésio Fernandes •
UNIVERSITY COLLEGE LONDON, REINO UNIDO, BRASIL.

Emerson Dionísio Gomes de Oliveira •
UNIVERSIDADE DE BRASÍLIA - UNB, BRASIL.

Estevam Barbosa de Las Casas •
UNIVERSIDADE FEDERAL DE MINAS GERAIS - UFMG, BRASIL.

Guilherme Ary Plonski •
UNIVERSIDADE DE SÃO PAULO - USP, BRASIL.

Heloisa Soares de Moura Costa •
UNIVERSIDADE FEDERAL DE MINAS GERAIS - UFMG, BRASIL.

Hugo Eduardo Araújo da Gama Cerqueira •
UNIVERSIDADE FEDERAL DE MINAS GERAIS - UFMG, BRASIL.

Joachim Michael •
UNIVERSIDADE DE BIELEFELD, ALEMANHA.

João Antônio de Paula •
UNIVERSIDADE FEDERAL DE MINAS GERAIS - UFMG, BRASIL.

José Antônio Rocha Gontijo •

UNIVERSIDADE ESTADUAL DE CAMPINAS - UNICAMP, BRASIL.

Leila Christina Dias •

UNIVERSIDADE FEDERAL DE SANTA CATARINA, UFSC - BRASIL

Luiz Bevilacqua •

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO, UFRJ - BRASIL.

Luiz Carlos Dias •

UNIVERSIDADE ESTADUAL DE CAMPINAS - UNICAMP, BRASIL.

Luiz Oosterbeek •

INTERNATIONAL COUNCIL OF PHILOSOPHY AND HUMAN SCIENCES, UNESCO.

Maria do Carmo de Freitas Veneroso •

UNIVERSIDADE FEDERAL DE MINAS GERAIS, UFMG, BRASIL.

Marília Andrés Ribeiro •

INSTITUTO MARIA HELENA ANDRÉS, BRASIL.

Ricardo Hiroshi Caldeira Takahashi •

UNIVERSIDADE FEDERAL DE MINAS GERAIS, UFMG, BRASIL.

Roberto Vecchi •

UNIVERSIDADE DE BOLONHA, ITÁLIA.

Sérgio Schneider •

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL - UFRGS, BRASIL

Comissão Editorial

Ana Maria Rabelo Gomes •

UNIVERSIDADE FEDERAL DE MINAS GERAIS - UFMG, BRASIL.

Estevam Barbosa de Las Casas •

UNIVERSIDADE FEDERAL DE MINAS GERAIS - UFMG, BRASIL.

Heloísa Soares de Moura Costa

UNIVERSIDADE FEDERAL DE MINAS GERAIS - UFMG, BRASIL.

Isabela Sales Prado •

UNIVERSIDADE FEDERAL DE MINAS GERAIS - UFMG, BRASIL.

Marília Andrés Ribeiro •

INSTITUTO MARIA HELENA ANDRÉS, BRASIL.

Rogério Palhares Zschaber de Araújo •

UNIVERSIDADE FEDERAL DE MINAS GERAIS - UFMG, BRASIL.

Pareceristas dessa edição

ANDRÉ GUILHERME COSTA-MARTINS

INSTITUTO DE PESQUISAS TECNOLÓGICAS - IPT, BRASIL.

ANDRÉ KASHIWABARA

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ - UTFPR, BRASIL.

ANNA PAULA RODRIGUES

UNIVERSIDADE FEDERAL DO TOCANTINS - UFT, BRASIL.

BRUNO RABELLO MONTEIRO

UNIVERSIDADE FEDERAL DE OURO PRETO - UFOP, BRASIL.

CARLOS FREDERICO DE BRITO D' ANDRÉA

UNIVERSIDADE FEDERAL DE MINAS GERAIS - UFMG, BRASIL.

CARMEN FLORES-MENDOZA

UNIVERSIDADE FEDERAL DE MINAS GERAIS - UFMG, BRASIL.

DANIEL DANTAS

UNIVERSIDADE FEDERAL DE SERGIPE - UFS, BRASIL.

DEMI GETSCHKO

UNIVERSIDADE DE SÃO PAULO - USP, BRASIL.

DIEGO COUTO DE LAS CASAS

UNIVERSIDADE FEDERAL DE MINAS GERAIS - UFMG, BRASIL.

EDEILSON MILHOMEM DA SILVA

UNIVERSIDADE FEDERAL DO TOCANTINS - UFT, BRASIL.

EDUARDO DA SILVA RIBEIRO

UNIVERSIDADE FEDERAL DE OURO PRETO - UFOP, BRASIL.

EULER HORTA MARINHO

UNIVERSIDADE FEDERAL DE OURO PRETO - UFOP, BRASIL.

EVANDRO LANDULFO TEIXEIRA PARADELA CUNHA

UNIVERSIDADE FEDERAL DE MINAS GERAIS - UFMG, BRASIL.

FABIO GAGLIARDI COZMAN

UNIVERSIDADE DE SÃO PAULO - USP, BRASIL.

FLAVIO VINICIUS DINIZ DE FIGUEIREDO

UNIVERSIDADE FEDERAL DE MINAS GERAIS - UFMG, BRASIL.

GREICE LACERDA

UNIVERSIDADE DO ESTADO DO RIO DE JANEIRO - UERJ, BRASIL.

GUSTAVO CARNIVALI

UNIVERSIDADE FEDERAL DE MINAS GERAIS - UFMG, BRASIL.

HELLIENE SOARES CARVALHO

INSTITUTO FEDERAL DO ESPÍRITO SANTO - IFES, BRASIL.

JOÃO RICARDO SATO

UNIVERSIDADE FEDERAL DO ABC - UFABC, BRASIL.

LEANDRO COELHO

UNIVERSIDADE FEDERAL DO PARANÁ - UFPR, BRASIL.

LEONARDO ALVARENGA LOPES SANTOS

UNIVERSIDADE FEDERAL DE SÃO JOÃO DEL-REI - UFSJ, BRASIL.

MARA MAGDA SOARES

UNIVERSIDADE FEDERAL DE MINAS GERAIS - UFMG, BRASIL.

MARCOS ALVES

UNIVERSIDADE FEDERAL DE MINAS GERAIS - UFMG, BRASIL.

MARCOS SEVERO

UNIVERSIDADE FEDERAL DE GOIÁS - UFG, BRASIL.

RICARDO HIROSHI CALDEIRA TAKAHASHI

UNIVERSIDADE FEDERAL DE MINAS GERAIS - UFMG, BRASIL.

RICARDO TOLEDO

UNIVERSIDADE FEDERAL DE SÃO JOÃO DEL-REI - UFSJ, BRASIL.

RODRIGO SILVA

UNIVERSIDADE FEDERAL DE OURO PRETO - UFOP, BRASIL.

SABRINA SEDLMAYER

UNIVERSIDADE FEDERAL DE MINAS GERAIS - UFMG, BRASIL.

SUZANA DE SIQUEIRA SANTOS

FUNDAÇÃO GETULIO VARGAS - FGV, BRASIL.

TAIANE COELHO RAMOS

UNIVERSIDADE FEDERAL FLUMINENSE - UFF, BRASIL.

VENÚNCIA COELHO

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE MINAS GERAIS - IFMG, BRASIL.

Revista da Universidade Federal de Minas
Gerais Universidade Federal de Minas
Gerais Av. Presidente Antônio Carlos,
nº 6.627, Campus Pampulha - Unidade
Administrativa III, Sala 151 - Pampulha -
Belo Horizonte - MG | CEP 31270-901 | +55
(31) 3409-4123 | 3409-5509

SUMÁRIO

09	Editorial
21	Sobre a artista e suas imagens ISABELA PRADO
25	Máquinas de produzir extra-naturezas: inteligência artificial e subversão botânica GISELLE BEIGUELMAN
32	O impacto da Inteligência artificial nas ciências da vida através da bioinformática LUCAS MORAES DOS SANTOS, DIEGO MARIANO E RAQUEL CARDOSO DE MELO-MINARDI
60	Infraestruturas, inteligência artificial e outras “tecnosoluções”: Google e a plataformização da emergência climática CARLOS D' ANDRÉA
80	A expansão da inteligência artificial e seu impacto nas dinâmicas sociais: desafios e responsabilidades JULIÃO GONÇALVES AMARAL
108	Inteligência artificial, tecnologia e as serventias extrajudiciais OSVALDO JOSÉ GONÇALVES DE MESQUITA FILHO E MARCELO DE OLIVEIRA MILAGRES
132	Imaginação e inteligência: apropriação, hibridismos, realidades e Inteligência artificial em uma exposição artística PABLO GOBIRA, ANA LUIZA PEDROSA CAMILO E EDUARDO SÉLLOS
168	Paralelos entre as inteligências natural e artificial a partir de um comparativo epistemológico: limitações e impactos LUIZ JOAQUIM NUNES E MARIANA INÉS GARBARINO
190	O potencial da inteligência artificial para a otimização do sistema de dimensionamento de conflitos OTÁVIO MORATO DE ANDRADE E DIERLE NUNES
218	Ética e inteligência artificial: desafios e melhores práticas CAROLINA DE MELO NUNES LOPES E JÚLIA CASTRO MENDES
246	Inteligência artificial, blockchain e a cadeia de custódia da prova no processo penal MATHEUS ARAÚJO
270	Monitoramento de integridade estrutural utilizando inteligência artificial: desafios, avanços e aplicações VICTOR HIGINO MENEGUITTE ALVES, VINICIUS ANTÔNIO MENEGUITTE ALVES E ALEXANDRE ABRAHÃO CURY

300

Preste atenção: a veloz evolução do processamento de linguagem natural e onde ela empaca
FÁBIO COZMAN E HUGO NERI

320

Proposta heurística para geração de soluções válidas para um problema de planejamento da produção florestal
ISABELLA ALKMIM BARRAL, LUCIANA BALIEIRO COSME E ALLYSSON STEVE MOTA LACERDA

350

Máquinas companheiras: tecnologias para o cuidado, autonomia e interdependência
PAULA LEMOS VILAÇA FARIA

366

A inteligência artificial representada em Genshin Impact, iniciativas regulatórias e letramento algorítmico
MAURÍCIO FERREIRA SANTANA E GERALDO MAGELA PIERONI

392

Superfície artificial, interior visceral: aspectos tecnológicos, estéticos e econômicos da arte generativa
RAFAEL GALVÃO DE ALMEIDA

420

A expansão artificial da inteligência humana
LUCIA SANTAELLA

442

Obstáculos para a uma mobilidade urbana autônoma: uma análise a partir da perspectiva humana
SANDOR BANYAI PEREIRA, IARA SOUSA CASTRO E RÓBER DIAS BOTELHO

472

Aprendizado de máquina e a previsão de curto prazo para velocidade máxima do vento
EDVANA DA SILVA DA SILVEIRA, POLÍBIO ÍCARO MORO CAPO E CARLA DE ABREU D'AQUINO

492

Estudo comparativo de modelos para previsão da mortalidade da SRAG por COVID-19 no Brasil
ANDRÉ VINÍCIUS MENDES BARROS, VICTOR HUGO DANTAS GUIMARÃES E LUCIANA BALIEIRO COSME

518

Reflexões epistemológicas e éticas sobre a inteligência artificial como área da computação cognitiva
MARIAH BROCHADO E LUCAS MAGNO DE OLIVEIRA PORTO

550

A vida real de Andrea, robô sem gênero
PARÉ ZAVEN

560

Sexismo e racismo algorítmico: um olhar interseccional sobre o neomaterialismo através do Midjourney
AMANDA MARIA DE SOBRAL GOMES

586

O pensamento não ocidental, a cosmotécnica de Yuk Hui e a tecnologia chinesa
BEATRIZ CARMO E SILVA, MARCELA VAROTTO MARAJÓ, MARCELO REIS MAIA E NATACHA RENA

612

Inteligência artificial e direito: o impacto transformador na organização de grandes acervos de textos jurídicos
DANIELA VIANNA, MARINA SOARES MARINHO, TIAGO FERREIRA E ALTIGRAN SOARES DA SILVA



INTELIGÊNCIA E INTELIGÊNCIA ARTIFICIAL

O XXI foi saudado como o século da biologia. Craig Venter e Daniel Cohen (2004) formularam essa ideia «Se o século 20 fosse o século da física, o século 21 seria o século da biologia. Enquanto a combustão, a eletricidade e a energia nuclear definiram o avanço científico no século passado, a nova biologia da pesquisa genômica – que fornecerá o projeto genético completo de uma espécie, incluindo a espécie humana – definirá a próxima». A biologia avançou tremendamente nas últimas duas décadas e teve uma grande participação no desenvolvimento da ciência e da indústria. No entanto, pode-se questionar se os recentes avanços da Inteligência Artificial, e seu potencial de transformação das práticas atuais na sociedade, não desafiam o suposto impacto certamente profundo das ciências biológicas como campo de pesquisa dominante. Evoluindo com altos e baixos, desde o século passado, a IA mostrou seu potencial para interferir na tomada de decisões e decisões em muitos aspectos da atividade humana. Desenvolvimentos recentes em linguagens naturais desafiam a velha percepção de que a substituição do trabalho humano pelas máquinas seria restrita a robôs que executam tarefas mecânicas inseguras ou tediosas. A recente pandemia trouxe mudanças nas práticas humanas que minimizam a necessidade de presença pessoal no local de trabalho, difundindo exponencialmente métodos de trabalho que estavam restritos a ambientes selecionados, proporcionando um impulso adicional às mudanças causadas pela IA.

Algumas questões éticas subjacentes envolvem o controle dos processos de tomada de decisão por algoritmos que são proprietários, projetados e do conhecimento restrito a um pequeno número de grandes conglomerados; a influência da nova divisão do trabalho entre nações e indivíduos, acelerando as tendências atuais de concentração e exclusão de hiper distribuição de riqueza; os riscos de atribuir decisões a máquinas e programas sem critérios humanos e muitas outras. Conforme Duarte (2019), os algoritmos de reconhecimento facial já conseguem identificar, por exemplo, todo o

grupo de quatro milhões de usuários do metrô em uma cidade como São Paulo, tão recentemente proposto como ferramenta de segurança (para «proteger» os usuários do sistema) e barrado após uma disputa judicial. Deve-se mencionar também que o conjunto de incertezas sobre o impacto da IA na sociedade inclui a possibilidade de que existam riscos mais profundos para a humanidade, ligados à manipulação direta por máquinas do próprio «sistema operacional» sobre o qual nossa civilização é construída: nossa linguagem. Exemplos em menor escala do tipo de dificuldades que possivelmente serão encontradas são os algoritmos das redes sociais, que vêm impactando de forma inesperada e ainda pouco compreendida diversos aspectos de nossa organização social. Note que esses algoritmos são muito mais simples do que a IA generalista que agora é anunciada.

O Instituto de Estudos Avançados de Paris e Belo Horizonte organizou um projeto para discutir e articular um grupo de pesquisadores denominado “Inteligência e Inteligência Artificial”. Devido à crise da Covid, o projeto foi prorrogado de 2020 para 2023. A relevância e a profundidade das discussões nos motivaram a propor e editar este número da Revista da UFMG, para o qual chamamos sua atenção.

O potencial de transformação e melhorias embarcadas em sistemas de IA mostra cada vez mais aplicações. Algoritmos de linguagem natural desenvolvidos recentemente trouxeram novas possibilidades surpreendentes, desde a elaboração de um laudo judicial até a resolução de um exame, avançando para a capacidade de escrever um programa de linguagem de computador para diferentes aplicações específicas. A revolução da IA faz parte de um contexto maior de tecnologias que promovem a transformação digital na economia mundial, na indústria e nos governos - essas tecnologias de transformação digital, incluindo a IA, apresentam riscos e oportunidades. O desafio atual para a sociedade é desenvolver um ambiente onde a IA possa ser uma ferramenta para melhorar a qualidade de vida, impactando energia, produção de alimentos, saúde e medicina, resposta a desastres, justiça e política.

A edição convidada sobre o tema Inteligência e Inteligência Artificial publica artigos para discutir esses impactos positivos e negativos, aspectos e aplicações da tecnologia de IA, e reflexões sobre todos os aspectos da inteligência natural e artificial. Os tópicos para este número são o desenvolvimento das técnicas hoje tratadas coletivamente como IA, suas interfaces com neurociências e ciências cognitivas, aplicações

transformadoras da IA em atividades econômicas, aspectos éticos e impactos em governança, trabalho e automação de tarefas cognitivas e intelectuais, limites de controle e liberdade, práticas governamentais, saúde e aplicações educacionais.

O processamento de linguagem natural é uma área da IA que lida com modelos de linguagem. Uma tarefa interessante é a modelagem de tópicos a partir de texto. Utilizando esta técnica, usamos uma IA para categorizar e agrupar os 25 artigos da edição de forma automática em áreas temáticas (Figura 1). O resultado foi o surgimento de quatro agrupamentos bem coerentes para os 25 trabalhos que se pode conferir neste número temático da Revista da UFMG:

Ética, filosofia e aspectos sociais

Arte generativa e artes digitais

Aspectos jurídicos e impactos práticos da IA\

Heurísticas e Aplicações em mobilidade, logística e indústria

Convidamos a uma proveitosa leitura dos artigos publicados nesta edição!

Registramos um agradecimento especial à aluna de doutorado em Ciência da Computação da UFMG, Giulia Zanon de Castro, por ter gerado os agrupamentos por meio de modelagem de tópicos com um modelo de linguagem LLM (Large Language Model).

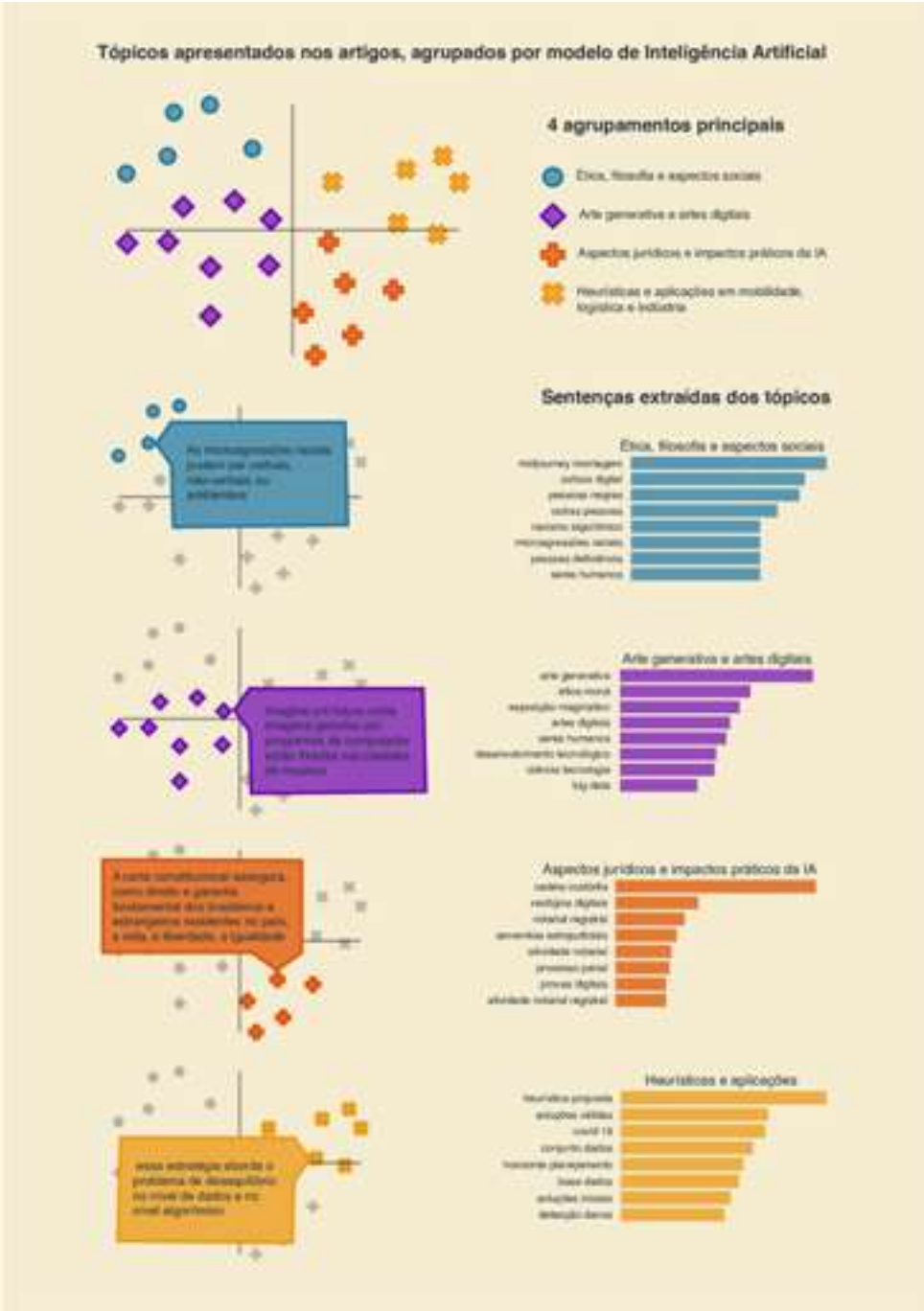
Uma ótima leitura a todos!

Frederico Gadelha Guimarães

Estevam Barbosa de Las Casas

André Fujita

Toshio Fukuda



Referências

DUARTE, Júlia Tupynambá. «A aplicação da tutela da proteção de dados pessoais no caso das portas interativas digitais do metrô de São Paulo.» *Trabalho de Conclusão de Curso (Bacharelado em Direito)-Faculdade Nacional de Direito, Universidade Federal do Rio de Janeiro, Rio de Janeiro (2019).*

VENTER, Craig, and Daniel Cohen. «The century of biology.» *New Persp. Q.* 21 (2004): 73.



INTELLIGENCE AND ARTIFICIAL INTELLIGENCE

The 21st has been hailed as the century of biology. Craig Venter and Daniel Cohen (2004) formulated this idea «If the 20th century were the century of physics, the 21st century would be the century of biology. While combustion, electricity and nuclear energy defined scientific advancement in the last century, The new biology of genomic research – which will provide the complete genetic blueprint for a species, including the human species – will define the next.” One may question whether recent advances in Artificial Intelligence, and their potential for transforming current practices in society, do not challenge the supposedly profound impact of biological sciences as a dominant field of research. Evolving with ups and downs since the last century. AI has shown its potential to interfere with decision-making and decisions in many aspects of human activity. Recent developments in natural languages challenge the old perception that the replacement of human labor by machines would be restricted to robots performing unsafe or tedious mechanical tasks. The recent pandemic has brought changes to human practices that minimize the need for personal presence in the workplace, exponentially spreading working methods that were previously restricted to select environments, providing an additional boost to the changes caused by AI.

Some underlying ethical issues involve the control of decision-making processes by algorithms that are proprietary, designed, and knowledge restricted to a small number of large conglomerates; the influence of the new division of labor between nations and individuals, accelerating current trends of concentration and exclusion and hyper-distribution of wealth; the risks of attributing decisions to machines and programs without human criteria and many others. According to Duarte (2019), facial recognition algorithms can already identify, for example, the entire group of four million subway users in a city like São Paulo, so recently proposed as a security tool (to «protect»

subway users). system) and blocked after a legal dispute. It should also be mentioned that the set of uncertainties about the impact of AI on society includes the possibility that there are deeper risks to humanity, linked to the direct manipulation by machines of the very «operating system» on which our civilization is built: our language. Smaller-scale examples of the type of difficulties that will likely be encountered are social media algorithms, which have had an unexpected and still little-understood impact on various aspects of our social organization. Note that these algorithms are much simpler than the generalist AI that is now advertised.

The Institute for Advanced Studies in Paris and Belo Horizonte organized a project to discuss and coordinate a group of researchers called “Intelligence and Artificial Intelligence”. Due to the Covid crisis, the project was extended from 2020 to 2023. The relevance and depth of the discussions motivated us to propose and edit this issue of the UFMG Magazine, to which we draw your attention.

The potential for transformation and improvements embedded in AI systems is showing more and more applications. Recently developed natural language algorithms have brought surprising new possibilities, from preparing a judicial report to solving an exam, advancing to the ability to write a computer language program for different specific applications. The AI revolution is part of a larger context of technologies that drive digital transformation in the global economy, industry and governments - these digital transformation technologies, including AI, present risks and opportunities. The current challenge for society is to develop an environment where AI can be a tool to improve quality of life, impacting energy, food production, health and medicine, disaster response, justice and politics.

The guest edition on the topic of Intelligence and Artificial Intelligence publishes articles to discuss these positive and negative impacts, aspects and applications of AI technology, and reflections on all aspects of natural and artificial intelligence. The topics for this issue are the development of techniques now collectively treated as AI, their interfaces with neurosciences and cognitive sciences, transformative applications of AI in economic activities, ethical aspects and impacts on governance, work and automation of cognitive and intellectual tasks, limits of control and freedom, government practices, health and educational applications.

Natural language processing is an area of AI that deals with language models. An

interesting task is topic modeling from text. Using this technique, we used AI to automatically categorize and group the 25 articles in the edition into thematic areas (Figure 1). The result was the emergence of four very coherent groupings for the 25 works that can be seen in this thematic issue of the UFMG Magazine:

Ethics, philosophy and social aspects

Generative art and digital arts

Legal aspects and practical impacts of AI

Heuristics and Applications in mobility, logistics and industry

We invite you to read the articles published in this edition!

We would like to express special thanks to the PhD student in Computer Science at UFMG, Giulia Zanon de Castro, for generating the clusters through topic modeling with an LLM (Large Language Model) language model.

A great read to everyone!

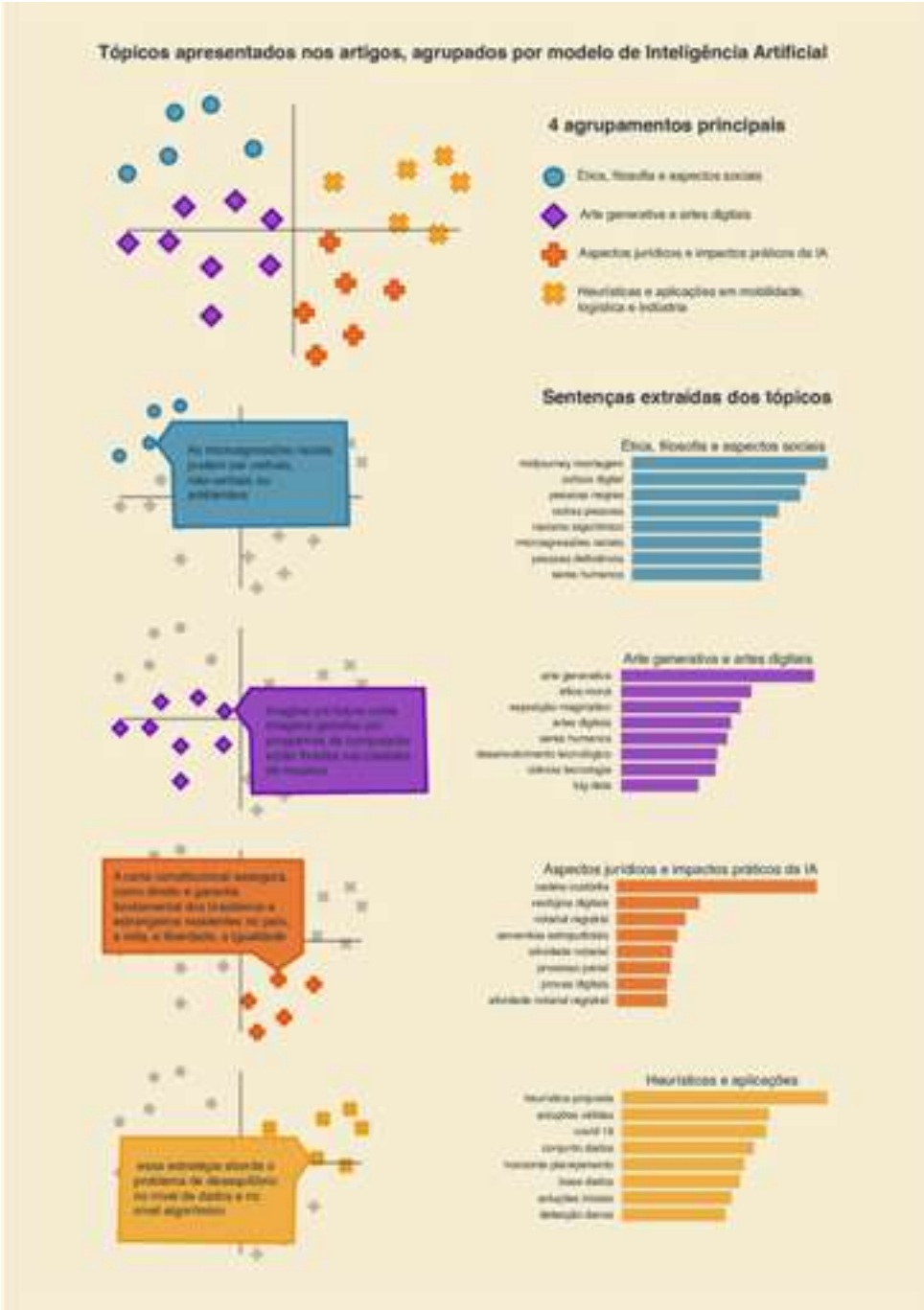
Frederico Gadelha Guimarães

Estevam Barbosa de Las Casas

André Fujita

Toshio Fukuda

1 A traditional expression. of brazilian literature containing rimed poems created from oral stories.



References

Duarte, Júlia Tupynambá. «A aplicação da tutela da proteção de dados pessoais no caso das portas interativas digitais do metrô de São Paulo.» *Trabalho de Conclusão de Curso (Bacharelado em Direito)-Faculdade Nacional de Direito, Universidade Federal do Rio de Janeiro, Rio de Janeiro* (2019).

Venter, Craig, and Daniel Cohen. «The century of biology.» *New Persp. Q.* 21 (2004): 73.



A SEMENTEIRA: O JARDIM EXTRA NATURAL DE GISELLE BEIGUELMAN

ISABELA PRADO*

Inteligência artificial é um dos grandes temas da contemporaneidade. Sua presença se manifesta em inúmeros campos e as consequências de sua expansão são imensuráveis e ainda não suficientemente compreendidas. Como um fenômeno social, a inteligência artificial reflete em boa medida os valores e ideologias característicos dos nossos tempos. Explorar esse tema tão vasto e tão atual é o desafio a que se dedica este número da Revista da UFMG. O ensaio visual que aqui se apresenta busca então contribuir para essa reflexão, de forma poética, considerando a abertura necessária ao se abordar um tema de tal relevância e abrangência.

Como em tantos outros campos, a inteligência artificial também se faz presente na produção artística. Mais do que uma mera ferramenta de geração de imagens ou outras formas de arte a partir do uso de algoritmos, a inserção da IA no campo das artes traz consigo todo um conjunto de implicações e questionamentos éticos, estéticos, processuais e mercadológicos.

A artista convidada para esta edição é Giselle Beiguelman, artista, pesquisadora e professora. O diálogo entre arte e tecnologia é tema longamente explorado em seus trabalhos, por meio dos quais Beiguelman reflete de forma crítica e inovadora sobre as relações entre estética e política, entre tecnologia e geração de imagens, entre cultura e sociedade. No que se refere ao uso de ferramentas de IA no campo das artes, Beiguelman é uma importante referência no cenário nacional e internacional. Seu trabalho instiga a reflexão em torno importantes questões políticas e sociais, tendo a IA como instrumento e como conceito.

* Artista visual e professora na Escola de Belas Artes da UFMG

Para este número da revista, foram selecionadas obras que compõem o corpo de trabalhos *Botannica Tirannica*, apresentada em museus no Brasil e no exterior. Neste caso, Beiguelman pesquisa centenas de espécies de plantas com nomes que denotam preconceito – machistas, racistas e antissemitas – e revela como a taxonomia botânica reflete uma visão de mundo preconceituosa e colonizadora, reproduzida pelos sistemas da Inteligência Artificial.

Flora Rebellis é uma série de vídeos produzidos com IA a partir de conjuntos de dados de plantas com nomes ofensivos e preconceituosos em relação a mulheres, negros, indígenas, judeus e “ciganos”. Fazendo um contraponto ao princípio da eugenia – que buscava identificar “padrões” entre inúmeros indivíduos de determinado grupo social – Beiguelman usa a Inteligência Artificial para procurar padrões internos entre as diferentes plantas dentro da mesma categoria (mulheres, negros, ciganos, etc.), mas impede o programa de chegar a um resultado uniforme, rebelando-se contra as suas próprias regras de funcionamento e criando “extra-naturezas”.

Venenosas, por sua vez, é uma série em que imagens são criadas por IA – com o uso da conversão de texto em imagem e de imagem em imagem – a partir da combinação de plantas ditas “carnívoras” com mandrágoras e beladonas. “*O foco está nas plantas proibidas pelo processo ‘civilizador’ colonial devido ao seu uso em rituais sagrados e poderes alucinógenos e afrodisíacos, contradizendo o cânone do ser humano ideal moral e fisicamente superior (homem europeu branco)*” (Beiguelman, 2023, p. 10, tradução livre).

O projeto *Venenosas* também é o foco do texto introdutório que Giselle Beiguelman apresenta para este número da revista. Este trabalho toma como referência mulheres apagadas da História da Arte e da Ciência. Ao identificar uma associação entre certas espécies de plantas, em sua maioria insetívoras, e ao papel das mulheres na sociedade, Beiguelman aponta para (e questiona) certos estereótipos sobre a feminilidade.

A introdução da inteligência artificial em várias áreas e vários contextos da vida social representa uma revolução cujas consequências ainda não foram totalmente exploradas. O campo das artes se apresenta como um local em que a IA também se faz presente de forma crescente, e ao mesmo tempo em que se constitui como um campo que propõe uma reflexão crítica sobre o uso das ferramentas de IA e seus impactos. Com trabalhos que utilizam a IA como instrumento para tratar de temas como colonialismo e misoginia, Giselle Beiguelman contribui de forma sensível para a investigação proposta neste número da Revista da UFMG.





MÁQUINAS DE PRODUZIR EXTRA-NATUREZAS:

inteligência artificial e subversão botânica

GISELLE BEIGUELMAN*

As tecnologias de inteligência artificial (IA) têm desempenhado papéis controversos. Por um lado, contribuem para a padronização e o controle social. Por outro lado, permitem desafiar os processos normativos, apontando para as potencialidades de uma “extra-natureza”, por meio de novas interlocuções entre humanos e máquinas, temas que explorei em algumas publicações recentes.

No momento, me dedico à exploração de modelos de inteligência artificial que utilizam linguagem natural para gerar imagens a partir de textos e transformar imagens em um dinâmico processo de tradução quase intersemiótica. As obras visuais apresentadas neste ensaio foram criadas com esse tipo de tecnologia, através de dois métodos complementares: conversão de texto em imagem e de imagem em imagem.

Elas são parte de um projeto em desenvolvimento, intitulado Venenosas, sobre plantas proibidas que foram proibidas ou demonizadas pelo colonialismo, que desenho com IA, tendo como referência mulheres apagadas da História da Arte e da Ciên-

* Giselle Beiguelman é artista e professora da FAU-USP. É autora de Políticas da imagem: vigilância e resistência na dadosfera (UBU Editora, 2021) e Memória da amnésia: políticas do esquecimento (Edições Sesc, 2019), entre outros. Suas obras artísticas integram acervos de museus no Brasil e no exterior, como ZKM (Alemanha), Jewish Museum Berlin, MAC-USP e Pinacoteca de São Paulo. Em seus projetos recentes investiga a construção do imaginário colonialista das artes e das ciências com recursos de Inteligência Artificial. É coordenadora do Projeto Temático Fapesp Acervos Digitais e Pesquisa: arte, arquitetura, design e tecnologia. Recebeu vários prêmios nacionais e internacionais. Site pessoal: <https://www.desvirtual.com/>

cia. O título provém de um livro do século XIX, publicado pela Sociedade Científica Cristã da Inglaterra (*Poisonous, Noxious, and Suspicious*). As razões das interdições dessas plantas variam muito, cobrindo um arco que vai do seu emprego em práticas rituais, suas funções afrodisíacas e sua difícil apreensão pelo imaginário rígido da taxonomia clássica, como as plantas ditas carnívoras.

Essas plantas, em sua grande maioria insetívoras e não carnívoras, têm intrigado o público por séculos, devido à sua peculiaridade em obter nutrientes. Ao invés de dependerem principalmente da fotossíntese, elas capturam e digerem pequenos animais, geralmente insetos, para suprir suas necessidades, muitas vezes em solos pobres em minerais. Entre as mais conhecidas estão as chamadas de armadilha de Vênus (*Dionaea muscipula*).

Seu nome científico faz referência a Dione, deusa das ninfas. Já o seu nome comum refere-se à deusa romana do amor e da beleza, Vênus, associando a planta tanto com a sensualidade quanto com a perigosidade — uma duplicidade que muitas vezes é imposta ao corpo feminino na literatura, na arte e nos meios de comunicação. A boca aberta da planta, que se fecha rapidamente sobre sua presa, foi frequentemente comparada a imagens de sedução e de perigo, evocando a “femme fatale”, uma figura que usa sua atratividade para enganar e prender.

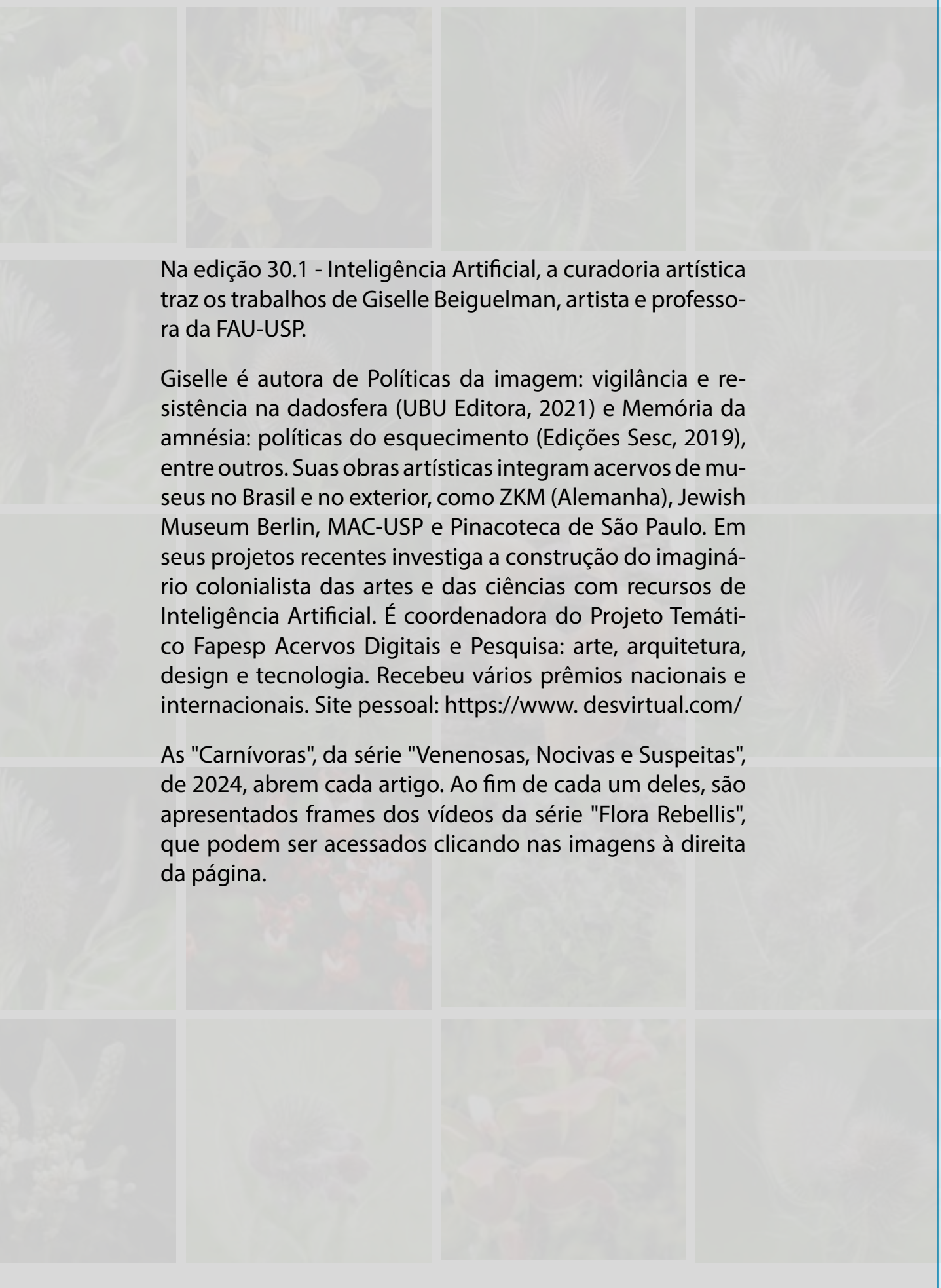
No século XIX, a era vitoriana forneceu terreno fértil para essa associação devido à sua atitude ambivalente em relação à sexualidade e ao papel das mulheres na sociedade. A planta carnívora, com suas formas sugestivas e método de nutrição ativo - considerado quase “predatório” - tornou-se uma metáfora para a preocupação vitoriana com a repressão e o desejo.

Essa associação também pode ser vista na arte e na cultura pop, na qual as plantas carnívoras são às vezes representadas como sedutoras e perigosas, refletindo e perpetuando estereótipos sobre a feminilidade. Filmes, como *A pequena loja dos horrores* (1986) e *O Senhor dos Anéis* (2001), e até alguns Pokémons, evocam imagens de uma “natureza consumidora” que está alinhada com arquétipos da suposta destrutividade das carnívoras.

No ensaio visual que apresento aqui, as carnívoras são o resultado do cruzamento de plantas ditas “carnívoras” com mandrágoras e belladonnas. São plantas que povoam os mais incríveis temores da cultura ocidental. Cercadas de preconceitos, assombram

pela sua beleza e são recorrentemente associadas à mulher e seus fictícios poderes de enfeitiçamento, captura e morte. Foram feitas com IA a partir de referências de várias mulheres botanistas que, no seu tempo, nunca foram reconhecidas, como Berthe Hoola van Noten (1817-1892), Sarah Anne Drake (1803-1857), que ilustrou o maravilhoso *Orchidaceae of Mexico and Guatemala* (1837-1843), de James Batemane, e Eliza e Sarah Maund, que ilustraram o fantástico *The Botanist* (1825-1846), de seu pai, Benjamin Maund.

Nestas imagens, busco não apenas resgatar essas mulheres e seu trabalho esquecidas, mas também provocar uma reflexão crítica sobre como interpretamos e interagimos com a natureza. Ao transformar digitalmente plantas historicamente marginalizadas em entidades visuais complexas, busco revelar camadas de significado que se acumularam em torno delas ao longo dos séculos. O objetivo é desconstruir os mitos e preconceitos enraizados que ainda hoje influenciam nossa percepção do mundo e da tecnologia.



Na edição 30.1 - Inteligência Artificial, a curadoria artística traz os trabalhos de Giselle Beiguelman, artista e professora da FAU-USP.

Giselle é autora de Políticas da imagem: vigilância e resistência na dadosfera (UBU Editora, 2021) e Memória da amnésia: políticas do esquecimento (Edições Sesc, 2019), entre outros. Suas obras artísticas integram acervos de museus no Brasil e no exterior, como ZKM (Alemanha), Jewish Museum Berlin, MAC-USP e Pinacoteca de São Paulo. Em seus projetos recentes investiga a construção do imaginário colonialista das artes e das ciências com recursos de Inteligência Artificial. É coordenadora do Projeto Temático Fapesp Acervos Digitais e Pesquisa: arte, arquitetura, design e tecnologia. Recebeu vários prêmios nacionais e internacionais. Site pessoal: <https://www.desvirtual.com/>

As "Carnívoras", da série "Venenosas, Nocivas e Suspeitas", de 2024, abrem cada artigo. Ao fim de cada um deles, são apresentados frames dos vídeos da série "Flora Rebellis", que podem ser acessados clicando nas imagens à direita da página.

 Giselle Beiguelman, frames do vídeo generativo da série Flora Rebellis, versão 2024.
Imagens geradas com Inteligência Artificial (Style Gans)

 Acesse o vídeo clicando na imagem









O IMPACTO DA INTELIGÊNCIA ARTIFICIAL NAS CIÊNCIAS DA VIDA ATRAVÉS DA BIOINFORMÁTICA

DOI: <https://doi.org/10.35699/2965-6931.2023.47996>

LUCAS MORAES DOS SANTOS*, DIEGO MARIANO**, RAQUEL CARDOSO DE MELO-MINARDI***

RESUMO: Nos últimos anos, as ciências da vida testemunharam uma profunda transformação, impulsionada pelos notáveis avanços na inteligência artificial (IA). Esta transformação tornou-se possível através da convergência de novos métodos e tecnologias que possibilitaram a geração de dados biológicos em larga escala e de alta qualidade. Além disso, a bioinformática tem desempenhado um papel fundamental viabilizando a modelagem e resolução de problemas biológicos complexos, criando assim um terreno fértil à aplicação do aprendizado de máquina e, consequentemente, catalisando insights e perspectivas inovadoras na área. Neste trabalho, discorremos sobre o profundo impacto da IA nas ciências da vida, com particular ênfase naqueles mediados pela bioinformática, na evolução contínua dos algoritmos de IA e nas implicações de longo alcance à pesquisa nas ciências da vida. Além disso, procuramos elucidar aspectos arquitetônicos inerentes às redes convolucionais e aos modelos generativos, demonstrando o motivo pelo qual cada técnica é aplicada a diferentes problemas biológicos.

PALAVRAS-CHAVE: bioinformática; inteligência artificial; aprendizagem de máquina.

The impact of Artificial Intelligence in life sciences through Bioinformatics

ABSTRACT En los últimos años, las ciencias de la vida han sido testigos de una profunda transformación, impulsada por avances notables en la inteligencia artificial (IA). Esta transformación fue posible gracias a la convergencia de nuevos métodos y tecnologías que permitieron la generación de datos biológicos a gran escala y de alta calidad. Además, la bioinformática ha jugado un papel fundamental al permitir el modelado y la resolución de problemas biológicos complejos, creando así un terreno fértil para la aplicación del aprendizaje automático y, en consecuencia, catalizando conocimientos y perspectivas innovadoras en el área. En este trabajo, analizamos el profundo impacto de la IA en las ciencias de la vida, con especial énfasis en aquellas mediadas por la bioinformática, la evolución continua de los algoritmos de IA y las implicaciones de gran alcance para la investigación en ciencias de la vida. Además, buscamos dilucidar aspectos arquitectónicos inherentes a las redes convolucionales y modelos generativos, demostrando por qué cada técnica se aplica a diferentes problemas biológicos.

KEYWORDS: bioinformática; inteligencia artificial; aprendizaje automático.

* Universidade Federal de Minas Gerais / ** Universidade Federal de Minas Gerais / *** Universidade Federal de Minas Gerais

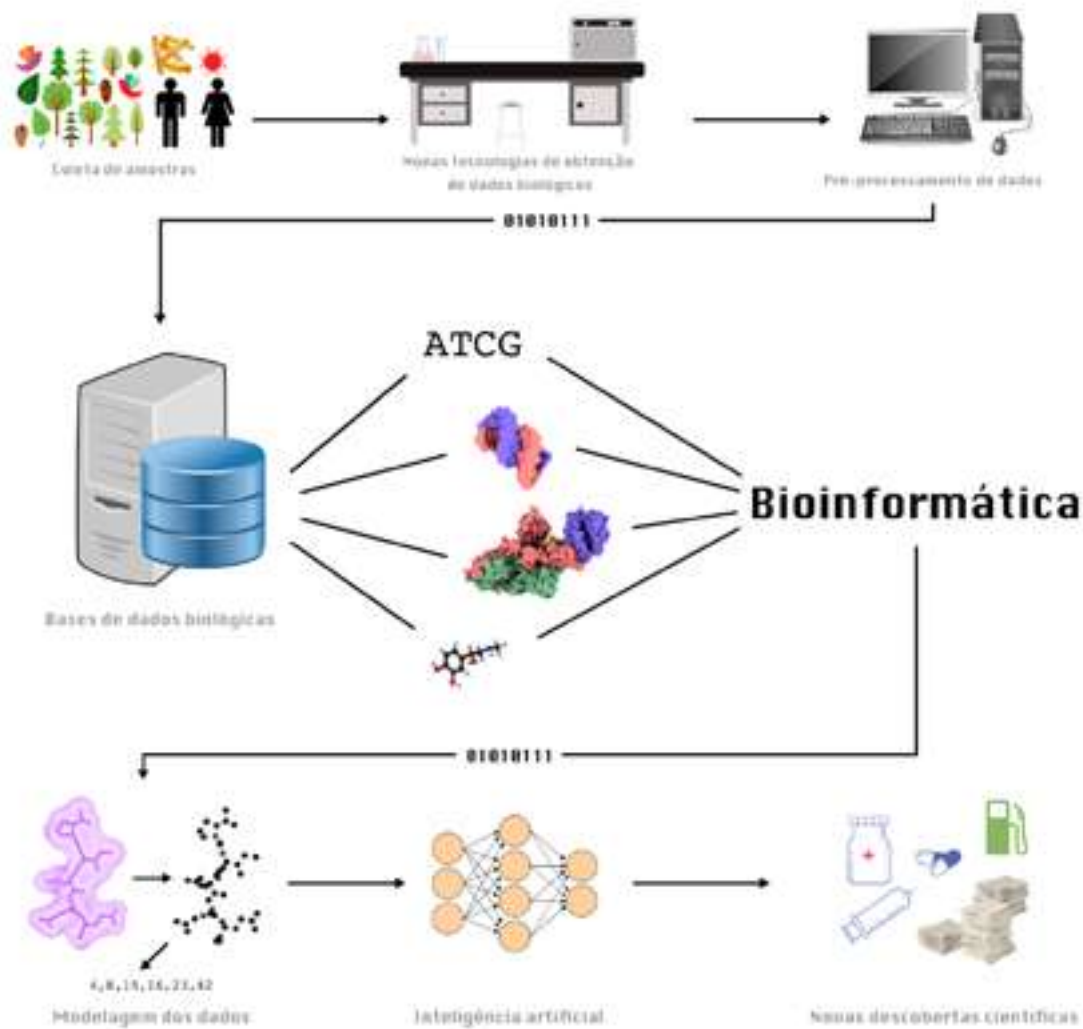
Introdução

Nos últimos anos, novas tecnologias, como sequenciadores de próxima geração ou equipamentos para realização de crio-microscopia eletrônica (Crio-ME), têm permitido a produção de uma imensa quantidade de dados biológicos. Além disso, diversas evoluções metodológicas e tecnológicas têm proporcionado um aumento expressivo do volume e da qualidade de bases de dados relacionados às biomoléculas. Nesse contexto, surge a bioinformática, uma área de pesquisa que envolve o uso e desenvolvimento de técnicas computacionais para análises de dados biológicos (Mariano, Ferreira, *et al.*, 2020; Mishra, Das, *et al.*, 2023).

A priori, estudos em bioinformática concentraram-se principalmente na análise de sequências de DNA, RNA e proteínas. No entanto, as pesquisas em bioinformática não se limitam a isso. Por exemplo, o interesse no estudo de estruturas tridimensionais de macromoléculas fomentou o surgimento de uma subárea conhecida como bioinformática estrutural. A bioinformática estrutural se baseia no uso de metodologias da química computacional em combinação com técnicas de modelagem molecular para lidar com questões biológicas de um ponto de vista tridimensional (Verli, 2014).

Recentemente, análises de bioinformática combinadas a algoritmos baseados em inteligência artificial (IA) têm possibilitado a pesquisadores obter novas percepções na abordagem de problemas por anos sem solução. A título de exemplo, o problema do enovelamento de proteínas, um desafio que havia sido estudado por mais de 50 anos (Dill e Maccallum, 2012), teve avanços significativos nos últimos anos com novas soluções envolvendo IA. Em 2021, pesquisadores do Google DeepMind propuseram uma nova arquitetura de redes neurais profundas denominada AlphaFold para predição de estruturas de proteínas a partir da estrutura primária de aminoácidos. Um artigo recente na revista *Nature* (Method of the Year 2021: Protein structure prediction, 2022) o indicou como “o método do ano de 2021” e previu ainda que seus resultados causariam um impacto duradouro e de longo alcance.

Figura 1 - Papel da bioinformática e das técnicas de inteligência artificial para obtenção de novas descobertas a partir de dados biológicos.



Fonte: autoria própria.

A Figura 1 apresenta um pipeline comum na pesquisa em bioinformática baseada em IA. Dado que a coleta de amostras de diversos organismos tem se tornado mais acessível, avanços recentes na tecnologia de sequenciamento de próxima geração (*Next-Generation Sequencing*, NGS) e técnicas para determinação de estruturas tridimensionais de proteínas por Crio-ME, têm possibilitado a obtenção rápida e eficiente de grandes volumes de informações biológicas (Gao, Mahajan, *et al.*, 2020). Em paralelo, o desenvolvimento da tecnologia da informação viabilizou uma capacidade ainda maior de processamento das bases de dados biológicas em larga escala. Além disso, metodologias de modelagem computacional têm permitido a preparação e representação dos dados (e.g., *one-hot encoding*, mapas de distâncias, grafos) para sua utilização em algoritmos de IA (Defresne, Barbe e Schiex, 2021). Assim, é possível aplicar técnicas de aprendizado de máquina, viabilizando a descoberta de conhecimento.

Nesse sentido, revisões acerca da aprendizagem profunda aplicada à bioinformática têm abordado diferentes pipelines que possibilitam da predição de estrutura até a geração de novo de sequências de proteínas. Em dos principais e predecessores trabalhos na área, Min et al. (2017) desenvolveram uma revisão abrangente, na qual descrevem sobre aplicações categorizadas por domínio na bioinformática (ômicas e processamento de imagens ou sinais biomédicos) e arquitetura de aprendizagem profunda (perceptron multicamadas, redes neurais convolucionais e recorrentes). Contudo, com o desenvolvimento da IA generativa, modelos generativos vêm sobrepondo arquiteturas como as redes recorrentes, na predição e modelagem de moléculas baseada em estrutura, sequência ou propriedades físico-químicas, aparecendo majoritariamente em revisões recentes (Huang, Boyken e Baker, 2016; Gao, Mahajan, *et al.*, 2020; Torrisi, Pollastri e Le, 2020; Bai, Liu, *et al.*, 2021; Defresne, Barbe e Schiex, 2021; Abbasi, Santos, *et al.*, 2022).

Neste trabalho, serão discutidos os impactos da IA na bioinformática e nas ciências da vida. Nas seções seguintes, serão apresentadas técnicas de aprendizagem profunda tradicionais como as redes neurais convolucionais, autoencoders variacionais, redes adversárias generativas e as redes neurais recorrentes, além de arquiteturas recentes como os modelos largos de linguagem (também conhecidos como modelos de linguagem baseados em transformer) e os modelos de difusão, dentre outros tópicos que serão descritos a partir de aplicações na bioinformática e ciências da vida.

Aprendizagem profunda

Nos últimos anos, o principal desafio à inteligência artificial envolveu a resolução de problemas que são inerentemente simples para humanos resolverem intuitivamente, mas que se tornam complexos quando é necessária uma descrição formal. Isso foi possível por meio de uma abordagem conhecida como hierarquia de conceitos que permite ao computador compreender conceitos complexos, construindo-os a partir de componentes mais elementares. Nesse sentido, ao relacionarmos esses conceitos através de um grafo, este será profundo (ou seja, composto por muitas camadas). Por esse motivo essa abordagem é conhecida como aprendizagem profunda em IA (Goodfellow, Bengio e Courville, 2016).

Em outras palavras, a solução dos problemas se dá pela extração de conhecimento (ou padrões não triviais) de bancos de dados através de algoritmos de aprendizado de máquina. No campo da Bioinformática, a IA tem sido empregada há mais de uma década impulsionada pelo crescente volume de dados biológicos advindos das ômicas (genômica, transcriptômica, proteômica, entre outras) ou ainda, de imagens e outros tipos de sinais biomédicos (Min, Lee e Yoon, 2017).

Contudo, o desempenho desses métodos ainda é dependente da representação dos dados (Goodfellow, Bengio e Courville, 2016). Para contornar esse desafio, tem sido utilizada uma abordagem alternativa – utilizar o aprendizado de máquina na descoberta de representações que podem ser expressas a partir de componentes mais simples (Duda, Hart e Stork, 2001). A necessidade de intervenção humana no processo de preparação dos dados pode ser significativamente diminuída com essa abordagem que é conhecida como aprendizado de representação (Bengio, Courville e Vincent, 2013). Dentre as diversas metodologias para aprender representações tem-se destacado a aprendizagem profunda (deep learning, DL), ou seja, um tipo especializado de aprendizado de máquina que alcança grande poder e flexibilidade aprendendo a representar o mundo a partir de uma hierarquia aninhada de conceitos (Goodfellow, Bengio e Courville, 2016).

Uma característica particular das arquiteturas baseadas em aprendizagem profunda é o uso de múltiplas camadas transformações não lineares, representadas por redes neurais, nas quais a saída de um nó é conectada a entrada de cada nó da camada seguinte (Bai, Liu et al., 2021). Esse arranjo das unidades neuronais forma camadas

ocultas e densas, constituindo uma rede totalmente conectada (Gonzalez e Woods, 2008, p. 945). Consequentemente, os neurônios ocultos possibilitam à rede extrair progressivamente as características mais significativas dos padrões de entrada, devido à profundidade das camadas (Haykin, 1999). Por esse motivo, essas redes neurais também são denominadas redes neurais profundas (deep neural networks, DNN).

Dessa forma, através de grandes volumes de dados sobre biomoléculas (DNA, RNA, proteínas, metabólitos, etc) e das redes neurais profundas, capazes de abstrair conceitos complexos compreendidos nos dados, muitos problemas têm sido resolvidos com maior precisão que se comparado a abordagens baseadas no desenvolvimento de programas clássicos, como fazia a bioinformática tradicionalmente.

Contudo, ainda que realizem previsões com precisão igual ou superior ao desempenho humano, a maioria dos modelos baseados em aprendizagem profunda são incapazes de fornecer resultados interpretáveis, frequentemente sendo comparados a uma “caixa preta” (Bai, Liu, et al., 2021). Isso ocorre porque as redes neurais profundas são capazes de mapear a entrada para representações internas de alta dimensionalidade, cujos padrões aprendidos encontram-se distribuídos nas camadas ocultas (Goodfellow, Bengio e Courville, 2016). Assim, torna-se difícil compreender diretamente algumas decisões específicas tomadas pela rede, limitando sua aplicação em sistemas biomoleculares (Li, Liu, et al., 2022). Visando contornar esse problema, um novo campo denominado aprendizado de máquina interpretável (interpretable machine learning, IML) tem alcançado interesse e popularidade. A partir do uso de técnicas como visualização de recursos, maximização de ativação e mecanismos de atenção, tornou-se possível obter insights sobre o que o modelo aprendeu (Molnar, 2022).

Vamos introduzir a seguir alguns dos principais tipos de redes neurais profundas que têm sido amplamente empregadas em bioinformática com significativa relevância.

Redes convolucionais

As redes neurais convolucionais (convolutional neural networks, CNN) são uma classe de rede neural profunda (LECUN, BOSER, et al., 1989), especializada no processamento de dados que possuem uma topologia em grade/matriz (Goodfellow, Ben-

gio e Courville, 2016, p. 330). Essas arquiteturas são inspiradas na hierarquia do córtex visual humano, na qual os neurônios iniciais respondem a padrões elementares em sub-regiões específicas de estímulos visuais, enquanto neurônios de nível superior sintetizam as informações dessas células mais simples, identificando padrões mais complexos (Min, Lee e Yoon, 2017).

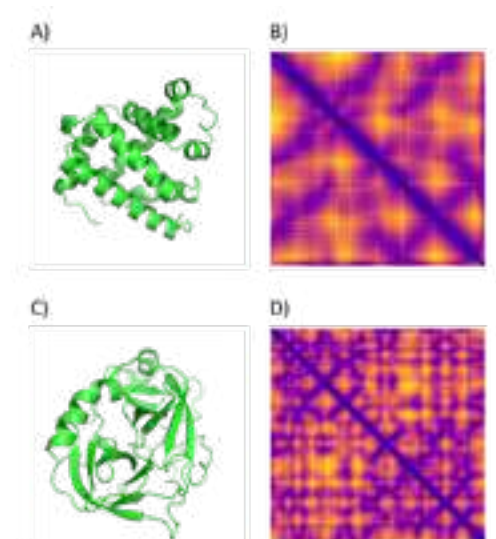
Em alusão ao seu nome, a CNN realiza uma operação linear denominada convolução que pode ser assimilada a um produto escalar de duas matrizes. Uma matriz contendo o conjunto de parâmetros aprendidos, conhecida como filtro, desliza pela imagem, convolvendo uma outra matriz que representa uma região local da entrada (Goodfellow, Bengio e Courville, 2016). Nesse caso, as camadas convolucionais não aprendem a entrada globalmente de uma única vez, mas concentram-se no padrão local do campo receptivo (Bian e Xie, 2021). Essa característica possibilita as CNNs aprenderem de padrões elementares (e.g., bordas, texturas) a características e conceitos abstratos (e.g., faces, objetos). Isso se destaca como um exemplo de como princípios neurocientíficos influenciam a aprendizagem profunda (Molnar, 2022).

No contexto da bioinformática estrutural, em que se estuda a estrutura 3D de proteínas e sua interação com outras biomoléculas, as CNNs têm sido amplamente utilizadas na predição de estrutura ou função de proteínas (Gao, Mahajan, et al., 2020; Torrisi, Pollastri e Le, 2020). Isto tem sido possível pois as CNNs se beneficiam das simetrias das representações estruturais em duas dimensões, como por exemplo dos mapas de distâncias que são representações em imagens das distâncias entre todos os pares de átomos de uma proteína (Defresne, Barbe e Schiex, 2021).

O exemplo mais notório é a primeira versão do AlphaFold¹ (Alquraishi, 2019; Senior, Evans, et al., 2020), cujo componente central é uma rede residual convolucional treinada com estruturas do PDB para prever distogramas¹ de proteínas, com base em sequências de aminoácidos. Além disso, as CNNs demonstraram ser relevantes na identificação dos estados funcionais de proteínas e peptídeos estáticos (Santos e Melo-Minardi, 2022; Santos et al., 2023), bem como daqueles simulados através da dinâmica molecular (Han, Lee, et al., 2022), usando uma representação por mapas de distâncias (Figura 2), ou pixels (Plante, Shore, et al., 2019; Li, Liu, et al., 2022), e na modelagem ab initio de proteínas (Anishchenko, Pellock, et al., 2021).

¹ Representação 2D na qual cada pixel corresponde distribuição de probabilidade do quanto próximo encontram-se os pares de resíduos no mapa de distâncias (SENIOR, EVANS, et al., 2020).

Figura 2 – Mapas de distâncias. A) Estrutura tridimensional de uma mioglobina (PDB ID: 1TES). B) Mapa de distâncias correspondente a mioglobina. C) Estrutura tridimensional de uma tripsina (PDB ID: 1SGT). D) Mapa de distâncias correspondente a tripsina.



Fonte: autoria própria.

Inteligência artificial generativa

Os modelos baseados em IA generativa, também conhecidos como *modelos generativos*, têm o objetivo de modelar uma distribuição a partir dos dados de treinamento e, posteriormente, produzir resultados semelhantes (Sanchez-Lengeling e Aspuru-Guzik, 2018; Zeng, Wang, *et al.*, 2022). Um exemplo bastante simples é treinar a rede com fotos de faces de seres humanos e depois utilizar o modelo generativo para produzir faces nunca vistas. Essa abordagem tem se mostrado promissora ao desenho de novas moléculas (inclusive proteínas), através da exploração não apenas do espaço de sequência, mas também de propriedades estruturais e funcionais (Huang, Boyken e Baker, 2016; Tong, Liu, *et al.*, 2021). Isso se deve ao fato de que cada molécula possui uma representação única em um espaço latente multidimensional, que codifica suas características. Por exemplo, considere uma molécula composta por três átomos A, B e C. Nesse caso, cada átomo poderia ser descrito por características, como seu elemento

químico (e.g., carbono, hidrogênio, oxigênio), carga (positiva, neutra e negativa), polaridade (apolar e polar) e suas coordenadas espaciais (x, y, z), por exemplo. Assim, a combinação dessas características resulta em potenciais moléculas em um espaço de possibilidades – referido como *espaço latente* – nesse caso 4D (elemento, carga, polaridade, posição).

Os modelos generativos exploram esse espaço variável, gerando vetores de características (ou latentes) a partir de uma distribuição de probabilidade (Sanchez-Lengeling e Aspuru-Guzik, 2018). Pesquisadores têm constatado que esses modelos treinados com sequências e estruturas de proteínas nativas, têm grande potencial na geração de moléculas *de novo* (do zero) (Callaway, 2023). Essas redes neurais têm transformado o desenho e a otimização de pequenas moléculas e macromoléculas, apresentando novos candidatos parcialmente otimizados, em alguns casos em um tempo inferior ao normalmente executado por abordagens convencionais (Zeng, Wang, et al., 2022; Thomas, Bender e Graaf, 2023). O desempenho desses modelos na química generativa se deve ao grande volume de dados disponíveis em bases de dados (Min, Lee e Yoon, 2017; Bian e Xie, 2021).

Autoencoders variacionais

Um *autoencoder* é uma rede neural treinada para tentar reconstruir a entrada em sua saída, com a maior precisão possível (Goodfellow, Bengio e Courville, 2016, p. 499). O aprendizado nessa arquitetura baseia-se em uma função, denominada codificadora, que transforma os dados em alta dimensionalidade (entrada) para uma representação de baixa dimensionalidade em um espaço latente, e um decodificador que reconstrói a entrada original a partir dessa representação ((Goodfellow, Bengio e Courville, 2016; Tong, Liu, et al., 2021). No entanto, no contexto de aplicações generativas, é mais frequente encontrar uma variante desse modelo conhecida como *autoencoder* variacional (*variational autoencoder*, Vae), sendo a entrada representada como uma distribuição de probabilidade no espaço latente (Kingma e Welling, 2013; Sousa, Correia, et al., 2021).

Essas arquiteturas têm sido amplamente empregadas no desenho de novas molé-

culas (Chenthamarakshan, Hoffman, *et al.*, 2023), uma vez que possibilitam a exploração de regiões desconhecidas do espaço de design químico a partir de representações latentes de baixa dimensionalidade (Anstine e Isayev, 2023). Nesse caso, o codificador mapeia a molécula em um *embedding* contínuo, ou seja, um vetor latente com dimensionalidade inferior à representação da entrada (Bian e Xie, 2021), enquanto o decodificador tenta recuperar a molécula a partir do *embedding* aprendido. Assim, o objetivo no treinamento de VAEs é minimizar o erro de reconstrução, ou seja, a diferença entre a representação de entrada fornecida ao codificador e a saída do decodificador (Anstine e Isayev, 2023).

Redes adversárias generativas

As redes adversárias generativas (*generative adversarial networks*, GAN) são arquiteturas de redes neurais profundas compostas por dois módulos concorrentes: gerador e discriminador (Goodfellow, Pouget-Abadie, *et al.*, 2014). O aprendizado em GANs pode ser compreendido como um jogo de soma zero, onde o gerador aprende a distribuição dos dados de treinamento criando instâncias sintéticas para classificação pelo discriminador. Conforme o treinamento avança, o gerador melhora, produzindo resultados cada vez mais realistas para enganar o discriminador e classificá-los como amostras autênticas. Simultaneamente, o discriminador se torna mais eficiente em distinguir instâncias reais das falsas. Eventualmente, as amostras do gerador tornam-se indistinguíveis dos dados reais, tornando o discriminador desnecessário à medida que o treinamento converge (Goodfellow, Bengio e Courville, 2016, p. 702), condição conhecida como *Equilíbrio de Nash* (Nash, 1950).

Esses modelos podem ser conceitualmente simples de compreender a partir de um exemplo de geração de moléculas. O discriminador visa maximizar a taxa de erro das moléculas sintéticas do gerador, enquanto este, por sua vez, tenta minimizar o erro a partir da criação de moléculas suficientemente realistas, capazes de enganar o discriminador (Bilodeau, Jin, *et al.*, 2022; Zeng, Wang, *et al.*, 2022). Ou seja, mesmo um discriminador bem treinado pode classificar instâncias geradas como reais, mostrando a capacidade do gerador em criar compostos promissores a partir de uma

entrada aleatória latente (Bian e XIE, 2021). Assim, esses modelos têm sido adaptados para lidar com diversos desafios na química generativa, abrangendo a criação de moléculas *de novo* (Anand e Huang, 2018; Surana, Arora, *et al.*, 2023; Xie, Valiente e Kim, 2023), predição da estrutura de proteínas (Ding e Gong, 2020) e modelagem de *loop* (Li, Nguyen, *et al.*, 2017).

Redes neurais recorrentes

As redes recorrentes (*recurrent neural networks*, RNN) são uma família de redes neurais para processamento de dados sequenciais (Rumelhart, Hinton e Williams, 1986; Goodfellow, Bengio e Courville, 2016). Essas arquiteturas apresentam *loops* internos, nos quais as saídas das camadas no estado anterior são transferidas para o estado atual (Mitchell, 1997). Dessa forma, elas conseguem processar a entrada consecutivamente, ao invés de uma única etapa (Bian e Xie, 2021). Contudo, o aumento das etapas na RNN pode ocasionar o desaparecimento de gradientes na retropropagação, dificultando o aprendizado de dependência a longo prazo. A inclusão de unidades LSTM (*Long Short-Term Memory*) contorna esse problema, introduzindo parâmetros aprendíveis para controlar o fluxo de informações (Sousa, Correia, *et al.*, 2021).

Essas arquiteturas têm sido empregadas em tarefas de modelagem de sistemas que possuem um componente sequencial, como a geração de moléculas a partir da sequência primária (Min, Lee e Yoon, 2017; Zeng, Wang, *et al.*, 2022). Neste contexto, é comum ser utilizada uma notação conhecida como Smiles (*Simplified Molecular Input Line Entry System*), ou seja, codificações textuais simplificadas, como *tokens*, para descrever estruturas moleculares, tornando-o uma escolha predominante para representar pequenas moléculas (Sousa, Correia, *et al.*, 2021). Após treinar a RNN com muitas strings Smiles, seria possível prever novas sequências válidas não presentes no conjunto de dados inicial (Tong, Liu, *et al.*, 2021). Por exemplo, a partir de uma sequência inicial (e.g., “CC”), a RNN atribui probabilidades a caracteres subsequentes. Neste caso, “I” (um) corresponderia a alta probabilidade, podendo ser escolhido como o próximo caractere. “I” seria a entrada de feedback à RNN. Este ciclo continua até o token final, “\n”, denotando o fim da sequência (Zeng, Wang, *et al.*, 2022).

Modelos largos de linguagem

Modelos largos de linguagem (*large language models*, LLM) são uma categoria especializada de modelo de linguagem caracterizada por uma rede neural profunda contendo um extenso número de parâmetros, comumente na ordem de 10^9 . Recentemente, eles tornaram-se o estado da arte no processamento de linguagem natural (Sousa, Correia, *et al.*, 2021) uma vez que utilizam uma arquitetura denominada *transformer* (Vaswani, Shazeer, *et al.*, 2017). Esse modelo consiste em uma estrutura codificador-decodificador, na qual um componente crucial, conhecido como mecanismo de atenção, permite a ele focar em partes relevantes da sentença. Além disso, devido ao paralelismo do *transformer*, as sentenças são processadas como um todo, prevenindo o desaparecimento do gradiente durante o treinamento (Zhang, Fan, *et al.*, 2023).

As similaridades intrínsecas às linguagens naturais e sequências biológicas têm impulsionado a aplicação de LLMs (e.g., Bert, GPT-3) na pesquisa em química e bioinformática, abrangendo da análise de sequência a modelagem *de novo* (Vert, 2023; Zhang, FAN, *et al.*, 2023). Isso, pois, dados biológicos heterogêneos e de alta dimensionalidade (e.g., descritores moleculares, estruturas de proteínas etc.) podem ser representados de maneira uniforme, através de *embeddings* (Tong, Liu, *et al.*, 2021; Vert, 2023). Um *embedding* é um vetor contínuo no espaço d -dimensional usado para codificar padrões complexos a partir dos dados de entrada (e.g., texto ou imagens). Eles são estruturados de forma a posicionar conceitos semelhantes próximos uns dos outros no espaço de *embeddings*, capturando relações semânticas. Esses vetores podem ser alimentados diretamente nos LLMs, que extraem informações dependentes do contexto biológico da correlação de segmentos de embeddings (Zhang, Fan, *et al.*, 2023).

Nesse sentido, tanto os *transformers* como os LLMs, têm sido o núcleo de métodos computacionais para predição de estruturas tridimensionais de proteínas com alta precisão, representando o estado da arte nesse campo. A partir da utilização de *embeddings* derivados de alinhamentos múltiplos de sequência e um modelo baseado em *transformer*, o AlphaFold 2 (Jumper, Evans, *et al.*, 2021) obteve previsões de estruturas de proteínas em paridade com métodos experimentais, alcançando uma pontuação GDT

de 92,4 no CASP14². Outro exemplo é o ESM-2³ (Lin, Akin, et al., 2023), uma nova família de modelos de linguagem de proteínas baseados em *transformers* pré-treinados generativos (generative pre-trained transformer, GPT) para inferir uma estrutura diretamente da sequência primária. Nesse caso, o ESM-2 aproveita informações de coevolução de resíduos em sequências para produzir previsões de nível atômico (Lin, Akin, et al., 2023; Zhang, Fan, et al., 2023).

Modelos de difusão

² .Disponível em: <https://github.com/RosettaCommons/RFdiffusion>.

Os modelos de difusão (*diffusion models*, Diff) (Sohl-Dickstein, Weiss, et al., 2015) têm sido bem-sucedidos na realização de tarefas como geração de imagens a partir de texto, produzindo amostras de alta qualidade (Anstine e Isayev, 2023). Essas redes têm como objetivo modelar uma distribuição de probabilidade condicional dos dados, a partir de etapas em sequência. Nesse caso, a estrutura de entrada (*e.g.*, imagem) é transformada gradualmente através de um procedimento iterativo conhecido como *difusão*. Esse processo envolve a introdução de um ruído gaussiano, que desloca os dados em direção a uma distribuição alvo, onde a estrutura resultante não possui qualquer semelhança com a estrutura inicial. Durante o treinamento, a rede aprende a remover o ruído a partir de uma estimativa da máxima verossimilhança, na qual o modelo visa maximizar a probabilidade de geração da entrada a partir dos dados ruidosos (Sohl-Dickstein, Weiss, et al., 2015).

Recentemente, uma equipe de pesquisadores desenvolveu um programa de IA denominado *RoseTTAFold diffusion*² (RFdiffusion) (Watson, Juergens, et al., 2023), que utiliza difusão guiada na criação de proteínas personalizadas (Callaway, 2023). Nesse caso, os autores ajustam a rede de predição de estrutura RoseTTAFold para o design de proteínas de acordo com restrições específicas de projeto durante a remoção de ruído processo conhecido como condicionamento (Callaway, 2023; Watson, Juergens, et al., 2023) alcançando excelente desempenho no design de backbones de proteínas. Além disso, modelos baseados em difusão têm sido apli-

cados na geração de conformação molecular, onde o modelo aprende a modificar iterativamente os pares de distâncias atômicas para gerar conformações estáveis (Bilodeau, Jin, *et al.*, 2022). Contudo, ainda que representem uma abordagem promissora, os modelos de difusão ainda são pouco explorados (Anstine e Isayev, 2023).

3 trRosetta.
Disponível em:
<https://yanglab.nankai.edu.cn/trRosetta/>.

Aprendizagem profunda aplicada a pesquisa em bioinformática

A diversidade de arquiteturas de aprendizagem profunda tem oferecido uma gama de alternativas à representação de dados biológicos, o que tem impulsionado o crescente e extenso volume de trabalhos em diferentes campos da bioinformática (Min Lee e Yoon, 2017). Além disso, programas de IA como o trRosetta³ (Du et al., 2021) e, mais recentemente, o AlphaFold2 (Jumper, Evans, et al., 2021) têm avançado para além da modelagem por homologia, sendo utilizados no design de novo de moléculas, e pavimentado o caminho para novas aplicações. Um exemplo é o AlphaFold Multimer, uma extensão do AlphaFold2 que foi desenvolvida especificamente para prever complexos proteína-proteína. Nesse sentido, a aprendizagem profunda e a IA generativa têm auxiliado no desenvolvimento ágil de moléculas personalizadas, explorando um espaço desconhecido de candidatos terapêuticos (Callaway, 2023). Com isso, torna-se possível a descoberta, ou otimização, de moléculas *in silico* para um determinado alvo, ou função, servindo de base para a descoberta de medicamentos e desenvolvimento de vacinas (Vert, 2023). O apoio computacional por meio da IA torna os processos de desenvolvimento de moléculas, tanto com fins biotecnológicos quanto farmacológicos, mais ágil, barato e promissor.

A tabela 1 apresenta uma síntese de trabalhos recentes em diferentes campos da bioinformática, onde as técnicas de aprendizagem profunda discutidas anteriormente, foram aplicadas em objetivos específicos com sucesso

Modelagem de proteínas <i>de novo</i>			
Arquitetura	Objetivo	Fonte	Repositório
CNN	Modelagem <i>de novo</i> de proteínas baseada em uma combinação de DNNs e simulações de refinamento estrutural, para gerar sequências alucinadas similares a sequências de proteínas nativas	ANISH-CHENKO, PELLLOCK, <i>et al.</i> , 2021	https://github.com/gjoni/trDesign
GAN	Modelagem generativa de estruturas de proteínas, codificadas como distâncias pareadas entre os carbonos no backbone da proteína	ANAND e HUANG, 2018	https://github.com/collinarneft/protein_gan
GAN	Geração e otimização de moléculas, usando uma representação de SMILES	ABBASI, SANTOS, <i>et al.</i> , 2022	https://github.com/larngroup/GAN-Drug-Generator
GAN	Geração de sequências para novos peptídeos antivirais	SURANA, ARORA, <i>et al.</i> , 2023	https://github.com/thoughworks/antiviral-peptide-predictions-using-gan/
GAN	Design <i>de novo</i> de -hélices representadas como vetores de sequência e <i>features</i> estruturais	XIE, VALIENTE e KIM, 2023	https://github.com/xxiexuezhi/helix_gan
Modelos de difusão	Geração de estruturas moleculares 3D condicionadas à forma de uma determinada molécula	CHEN, PENG, <i>et al.</i> , 2023	https://github.com/ehoogeboom/e3_diffusion_for_molecules
Modelos de linguagem baseados em <i>transformer</i>	Um modelo baseado em GPT pré-treinado para geração de sequências semelhantes a proteínas naturais a partir “do zero”	FERRUZ, SCHMIDT e HÖCKER, 2022	https://huggingface.co/nferruz/ProtGPT2
RNN	Descoberta de potenciais inibidores de quinase. As moléculas foram representadas por SMILES	LI, XU, <i>et al.</i> , 2020	https://github.com/Xyqii/RNN_generator
<i>Transformer</i>	Design racional de peptídeos antimicrobianos	MAO, GUAN, <i>et al.</i> , 2023	https://github.com/AspirinCode/AMPTrans-lstm

VAE	Design de inibidores para o domínio ligante ao receptor (RBD ⁴) da proteína spike e a principal protease do SARS-CoV-2 (M ^{pro})	CHENTHAMARAKSHAN, HOFFMAN, <i>et al.</i> , 2023	https://zenodo.org/records/7863805
Previsão de estrutura e função			
Arquitetura	Objetivo	Fonte	Repositório
CNN	Classificação de estados funcionais específicos a ligantes de GPCRs, usando conformações moleculares codificadas em trajetórias de DM transformadas em representações de pixel	PLANTE, SHORE, <i>et al.</i> , 2019	-
CNN	AlphaFold 1	ALQURAISHI, 2019; SENIOR, EVANS, <i>et al.</i> , 2020	https://github.com/google-deepmind/alphafold
CNN	Modelo DL interpretável baseado em representação por pixels para identificação de estados funcionais em trajetórias de DM de GPCRs, revelando resíduos-chave para compreensão do mecanismo funcional	LI, LIU, <i>et al.</i> , 2022	https://github.com/Jane-Liu97/ICNNMD
CNN	Reconhecimento de mudanças conformacionais relacionadas a padrões espaço-temporais induzidos por ligantes de receptores 2-AR usando mapas de distâncias	HAN, LEE, <i>et al.</i> , 2022	https://github.com/MinwooHan84/beta2AR
GAN	Modelagem de <i>loop</i> ⁵ usando uma representação de mapas de distâncias dos carbonos- da proteína	LI, NGUYEN, <i>et al.</i> , 2017	-
GAN	Previsão das distâncias inter-resíduos para proteínas, a partir de informações de coevolução 1D e 2D	DING e GONG, 2020	https://github.com/Wenze-Codibase/DistancePrediction-Protein-GAN.git
Modelos de linguagem baseados em <i>transformer</i>	ESM-1b	RIVES, MEIER, <i>et al.</i> , 2019; RAO, MEIER, <i>et al.</i> , 2020	https://github.com/facebookresearch/esm

Modelos de linguagem baseados em <i>transformer</i>	Modelo de linguagem profunda auto-supervisionado projetado especificamente para proteínas, o qual captura representações locais e globais de proteínas de maneira natural	BRANDES, OFER, <i>et al.</i> , 2022	https://github.com/nadavbra/protein_bert
Modelos de linguagem baseados em <i>transformer</i>	ESM-2	LIN, AKIN, <i>et al.</i> , 2023	https://github.com/facebookresearch/esm
<i>Transformer</i>	AlphaFold 2	JUMPER, EVANS, <i>et al.</i> , 2021	https://github.com/google-deepmind/alphafold
Expressão gênica			
Arquitetura	Objetivo	Fonte	Repositório
VAE	Geração de novas moléculas a partir de uma entrada constituída por compostos representados como SMILES e assinaturas de expressão gênica	PRAVAL-PHRUEKUL, PIRIYAJI-TAKONKIJ, <i>et al.</i> , 2023	https://github.com/ChemEXL/BiCEV
Modelos de linguagem baseados em <i>transformer</i>	Previsão da patogenicidade de variantes <i>missense</i> no genoma humano	BRANDES, GOLDMAN, <i>et al.</i> , 2023	https://github.com/ntranoslab/esm-variants

Tabela 1 – Síntese de aplicações representativas baseadas em aprendizagem profunda e IA generativa nos diferentes campos da pesquisa em bioinformática

Fonte: autoria própria.

4 Receptor Binding Domain

5 Predição de regiões ausentes em uma estrutura de proteína.

Considerações finais

Ao longo deste trabalho, discorremos sobre as principais técnicas de um subcampo da IA, conhecido como aprendizagem profunda, que têm sido amplamente adotadas na pesquisa em bioinformática. Buscamos também elucidar aspectos arquiteturais específicos às redes convolucionais e aos modelos generativos, demonstrando o motivo pelo qual cada técnica é aplicada em diferentes problemas biológicos.

A inteligência artificial tem um grande potencial para transformar a pesquisa em biologia e áreas afins. Estas áreas de pesquisa têm sido totalmente transformadas pelo uso de modelos computacionais nos últimos anos. Há um volume de dados crescente sobre os sistemas biológicos e que cresce também em qualidade e complexidade, de forma que tiram proveito dos novos modelos e algoritmos desenvolvidos pela bioinformática. A aplicação, ainda incipiente, de técnicas de IA nas ciências da vida por meio da bioinformática já trazem impactos consideráveis em diversas áreas de pesquisa e na computação, com a demanda por novos modelos e algoritmos. O que se vislumbra hoje é apenas o começo de um futuro brilhante desta parceria entre as ciências da vida e a ciência da computação.

Referências

- ABBASI, Maryam et al. *Designing optimized drug candidates with Generative Adversarial Network*. *Journal of Cheminformatics*, 14, n. 1, 2022. 40.
- ALQURAISHI, Mohammed. *AlphaFold at CASP13*. *Bioinformatics*, 35, n. 22, 2019. 4862-4865.
- ANAND, Namrata; HUANG, Po-Ssu. *Generative modeling for protein structures*. *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, 31, 2018. 7505-7516.
- ANISHCHENKO, Ivan. et al. *De novo protein design by deep network hallucination*. *Nature*, 600, n. 7889, 2021. 547-552.
- ANSTINE, Dylan M.; ISAYEV, Olexandr. *Generative Models as an Emerging Paradigm in the Chemical Sciences*. *Journal of the American Chemical Society*, 145, n. 16, 2023. 8736-8750.
- BAI, Qifeng et al. *Application advances of deep learning methods for de novo drug design and molecular dynamics simulation*. *WIREs Computational Molecular Science*, 12, n. 3, 2021. e1581.
- BENGIO, Yoshua; COURVILLE, Aaron; VINCENT, Pascal. *Representation Learning: A Review and New Perspectives*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35, n. 8, 2013. 1798-1828.
- BIAN, Yuemin; XIE, Xiang-Qun. *Generative chemistry: drug discovery with deep learning generative models*. *Journal of Molecular Modeling*, 27, n. 3, 2021. 71-89.
- BILODEAU, Camille et al. *Generative models for molecular discovery: Recent advances and challenges*. *WIREs Computational Molecular Science*, 12, n. 5, 2022.
- BRANDES, Nadav et al. *ProteinBERT: a universal deep-learning model of protein sequence and function*. *Bioinformatics*, 38, n. 8, 2022. 2102-2110.
- BRANDES, Nadav et al. *Genome-wide prediction of disease variant effects with a deep protein language model*. *Nature Genetics*, 2023.

CALLAWAY, Ewen. AI tools are designing entirely new proteins that could transform medicine. *Nature*, 619, n. 7969, 2023. 236-238.

CHEN, Ziqi et al. Shape-conditioned 3D Molecule Generation via Equivariant Diffusion Models. Preprint, 2023. Disponível em: <<https://arxiv.org/abs/2308.11890>>. Acesso em: 5 Setembro 2023.

CHENTHAMARAKSHAN, Vijil et al. Accelerating drug target inhibitor discovery with a deep generative foundation model. *Science Advances*, 9, n. 25, 2023.

DEFRESNE, Marianne; BARBE, Sophie; SCHIEX, Thomas. Protein Design with Deep Learning. *International Journal of Molecular Sciences*, 22, n. 21, 2021. 11741.

DILL, Ken A.; MACCALLUM, Justin L. The Protein-Folding Problem, 50 Years On. *Science*, 338, n. 6110, 2012. 1042-1046.

DING, Wenzhe; GONG, Haipeng. Predicting the Real-Valued Inter-Residue Distances for Proteins. *Advanced Science*, 7, n. 19, 2020. 2001314.

DU, Zongyang et al. The trRosetta server for fast and accurate protein structure prediction. *Nature Protocols*, 16, n. 12, 2021. 5634-5651.

DUDA, Richard O.; HART, Peter E.; STORK, David G. *Pattern Classification*. 2^{ff}. ed.

FERRUZ, Noelia; SCHMIDT, Steffen; HÖCKER, Birte. ProtGPT2 is a deep unsupervised language. *Nature Communications*, 13, n. 1, 2022. 4348.

GAO, Wenhao et al. Deep Learning in Protein Structural Modeling and Design. *Patterns*, 1, n. 9, 2020. 100142.

GONZALEZ, Rafael C.; WOODS, Richard E. *Digital Image Processing*. 4^{ff}. ed.

GOODFELLOW, Ian J. et al. Generative Adversarial Networks. *Advances in Neural Information Processing System*, 2014. 2672-2680.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. *Deep Learning*.

HAN, Minwoo et al. Recognition of the ligand-induced spatiotemporal residue pair pattern of β_2 -adrenergic receptors using 3-D residual networks trained by the time series of protein distance maps. *Computational and Structural Biotechnology Journal*, 20, 2022. 6360-6374.

HAYKIN, Simon. *Neural Networks: A Comprehensive Foundation*. 2^{ff}. ed.

HOLT, Charles A.; ROTH, Alvin E. The Nash equilibrium: A perspective. *Proceedings of the National Academy of Sciences*, 101, n. 12, 2004. 3999-4002.

HOPFIELD, J. J. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79, n. 8, 1982. 2554-2558.

HUANG, Po-Ssu; BOYKEN, Scott E.; BAKER, David. The coming of age of de novo protein design. *Nature*, 537, n. 7620, 2016. 320-327.

JUMPER, John et al. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596, n. 7873, 2021. 583-589.

KINGMA, Diederik P.; WELING, Max. *Auto-Encoding Variational Bayes*, 2013.

LECUN, Yan et al. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation*, 1, n. 4, 1989. 541-551.

LI, Chuan et al. An Interpretable Convolutional Neural Network Framework for Analyzing Molecular Dynamics Trajectories: a Case Study on Functional States for G-Protein-Coupled Receptors. *Journal of Chemical Information and Modeling*, 62, n. 6, 2022. 1399-1410.

LI, Xuanyi et al. Chemical space exploration based on recurrent neural networks: applications in discovering kinase inhibitors. *Journal of Cheminformatics*, 12, n. 1, 2020. 42.

LI, Yuesen et al. DrugGPT: A GPT-based Strategy for Designing Potential Ligands Targeting Specific Proteins. Preprint, 2023. Disponível em: <<https://doi.org/10.1101/2023.06.29.543848>>. Acesso em: 5 Setembro 2023.

LI, Zhaoyu et al. Protein Loop Modeling Using Deep Generative Adversarial Network. *IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI)*, 2017. 1085-1091.

LIN, Zeming et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379, n. 6637, 2023. 1123-1130.

MAO, Jiashun et al. Application of a deep generative model produces novel and diverse functional peptides against microbial resistance. *Computational and Structural Biotechnology Journal*, 21, 2023. 463-471.

MARIANO, Diego et al. A Brief History of Bioinformatics Told by Data Visualization. *Advances in Bioinformatics and Computational Biology. BSB 2020. Lecture Notes in Computer Science. Belo Horizonte: Springer, Cham*. 2020. p. 235-246.

METHOD of the Year 2021: Protein structure prediction. *Nature Methods*, 19, n. 1, 2022.

MIN, Seonwoo; LEE, Byunghan; YOON, Sungroh. Deep learning in bioinformatics. *Briefings in Bioinformatics*, 18, n. 5, 2017. 851-869.

MISHRA, Sarbani et al. Introduction to the World of Bioinformatics. *A Guide to Applied Machine Learning for Biologists*, 2023. 105-126.

MITCHELL, Tom M. *Machine learning*.

MOLNAR, Christoph. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. 2^{ff}. ed.

NASH, John F. Equilibrium points in *n*-person games. *Proceedings of the National Academy of Sciences*, 36, n. 1, 1950. 48-49.

PLANTE, Ambrose et al. A Machine Learning Approach for the Discovery of

Ligand-Specific Functional Mechanisms of GPCRs. *Molecules*, 24, n. 11, 2019. 2097.

PRAVALPHRUEKUL, Nutaya et al. De Novo Design of Molecules with Multi-action Potential from Diffe-

rential Gene Expression using Variational Autoencoder. *Journal of Chemical Information and Modeling*, 63, n. 13, 2023. 3999-4011.

RAO, Roshan. et al. Transformer protein language models are unsupervised structure learners. *International Conference on Learning Representations*, 2020.

RIVES, Alexander. et al. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *PNAS*, 118, n. 15, 2019.

ROSSETTO, Allison; ZHOU, Wenjin. GANDALF: Peptide Generation for Drug Design Using Sequential and Structural Generative Adversarial Networks. *Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, 2020.

RUMELHART, David E.; HINTON, Geoffrey E.; WILLIAMS, Ronald J. Learning representations by back-propagating errors. *Nature*, 323, n. 6088, 1986. 533-536.

SANCHEZ-LENGELING, Benjamin; ASPURU-GUZIK, Alán. Inverse molecular design using machine learning: Generative models for matter engineering. *Science*, 361, n. 6400, 2018. 360-365.

SANTOS, Lucas M. D. et al. Peptide-Protein Interface Classification Using Convolutional Neural Networks. *Advances in Bioinformatics and Computational Biology. BSB 2023. Lecture Notes in Computer Science*. Curitiba: Springer, Cham.

2023. p. 112-122.

SANTOS, Lucas M. D.; MELO-MINARDI, Raquel C. D. Identifying Large Scale Conformational Changes in Proteins Through Distance Maps and Convolutional Networks. *Advances in Bioinformatics and Computational Biology. BSB 2022. Lecture Notes in Computer Science*. Búzios: Springer, Cham. 2022. p. 56-67.

SENIOR, Andrew W. et al. Improved protein structure prediction using potentials from deep learning. *Nature*, 577, n. 7792, 2020. 706-710.

SOHL-DICKSTEIN, Jascha et al. Deep Unsupervised Learning using Nonequilibrium Thermodynamics. *ICML'15: Proceedings of the 32nd International Conference on International Conference on Machine Learning*. Lille: [s.n.]. 2015. p. 2256-2265.

SOUSA, Tiago et al. Generative Deep Learning for Targeted Compound Design.

Journal of Chemical Information and Modeling, 61, n. 11, 2021. 5343-5361.

SURANA, Shraddha. et al. PandoraGAN: Generating antiviral peptides. *SN COMPUT. SCI.*, 4, n. 607, 2023.

THOMAS, Morgan; BENDER, Andreas; GRAAF, Chris D. Integrating structure-based approaches in generative molecular design. *Current Opinion in Structural Biology*, 79, 2023. 102559.

TONG, Xiaochu et al. Generative Models for De Novo Drug Design. *Journal of Medicinal Chemistry*, 64, n. 19, 2021. 14011-14027.

TORRISI, Mirko; POLLASTRI, Gianluca; LE, Quan. Deep learning methods in protein structure prediction. *Computational and Structural Biotechnology Journal*, 18, 2020. 1301-1310.

VASWANI, Ashish et al. Attention is All you Need. *Advances in Neural Information Processing Systems*, 30, 2017. 6000-6010.

VERLI, Hugo. *Bioinformática: da biologia à flexibilidade molecular*.

VERT, Jean-Philippe. How will generative AI disrupt data science in drug discovery?. *Nature Biotechnology*, 41, n. 6, 2023. 750-751.

WATSON, Joseph L. et al. De novo design of protein structure and function with RFdiffusion. *Nature*, 620, n. 7976, 2023. 1089-1100.

XIE, Xuezhhi; VALIENTE, Pedro A.; KIM, Philip M. HelixGAN a deep-learning methodology for conditional de novo design of β -helix structures. *Bioinformatics*, 39, n. 1, 2023.

ZENG, Xiangxiang et al. Deep generative molecular design reshapes drug discovery. *Cell Reports Medicine*, 3, n. 12, 2022. 100794.

ZHANG, Shuang et al. Applications of transformer-based language models in bioinformatics: a survey. *Bioinformatics Advances*, 3, n. 1, 2023.





■ Giselle Beiguelman, frames do vídeo generativo da série Flora Rebellis, versão 2024.
Imagens geradas com Inteligência Artificial (Style Gans)

■ [Acesse o vídeo clicando na imagem](#)





INFRAESTRUTURAS, INTELIGÊNCIA ARTIFICIAL E OUTRAS “TECNOSOLUÇÕES”:

Google e a plataformação da emergência climática

DOI: <https://doi.org/10.35699/2965-6931.2023.47985>

CARLOS D'ANDRÉA*

RESUMO: O objetivo principal deste estudo é discutir a crescente vinculação entre plataformas infraestruturais e a crise climática em curso. Chamamos de “plataformização da emergência climática” os modos como as lógicas computacionais, econômicas e ideológicas de Big Techs como a Google incidem sobre desafios socioambientais contemporâneos. A partir de um estudo exploratório das ações de “sustentabilidade” da Google Brasil entre 2020 e 2023, problematizamos o duplo - e paradoxal - papel das plataformas infraestruturais: ao mesmo tempo em que se posicionam responsáveis pela produção de conhecimentos baseados em Inteligência Artificial voltados para mitigação dos desafios climáticos, contribuem para um maior consumo de energia, água e outros recursos naturais, sobretudo através seus data centers. Para subsidiar essa discussão, nos ancoramos nos Estudos de Plataforma e de Infraestruturas, que singularizam pela abordagem crítica sobre as disputas de poder se organizam a partir de materialidades.

RESUMO: plataformação; infraestrutura; inteligência artificial; emergência climática; Google.

Infrastructures, Artificial Intelligence and other “techno-solutions”: Google and the Platformization of the Climate Emergency

ABSTRACT: The main objective of this study is to discuss the growing link between infrastructural platforms and the on-going climate crisis. We call “platformization of the climate emergency” the ways in which the computational, economic and ideological logics of Big Techs like Google affect contemporary socio-environmental challenges. Based on an exploratory study of Google Brazil's “sustainability” actions between 2020 and 2023, we problematize the double - and paradoxical - role of infrastructural platforms: at the same time that they are responsible for the production of knowledge based on Artificial Intelligence aimed at mitigating climate challenges, reducing greater consumption of energy, water and other natural resources, mainly through its data centers. To support this discussion, we anchor ourselves in Platform and Infrastructure Studies, which stand out for their critical approach to how power disputes are organized based on materialities.

KEYWORDS: platformization. Infrastructure. artificial intelligence. climate emergency. Google.

* Universidade Federal de Minas Gerais

Introdução

Desinformação, discursos de ódio, uberização do trabalho, privacidade, bolhas ideológicas etc.: não são poucos os desafios sociopolíticos contemporâneos diretamente associados às lógicas e às estratégias comerciais de plataformas setoriais (como a Uber), mídias sociais (TikTok, Instagram etc) ou das plataformas infraestruturais (em especial, Amazon, Meta e Google). Em meio às muitas controvérsias e incertezas das quais as plataformas são protagonistas, uma parece não ganhar o merecido destaque: suas relações com o agravamento da emergência climática. Se no campo científico e mesmo na política não há dúvidas quanto à intensidade e urgência da crise climática¹, menos visíveis e debatidas são suas ligações com o acelerado processo de plataformização de inúmeras atividades econômicas e práticas sociais (Poell *et al*, 2020; Cristofari, 2023).

¹ Algumas conclusões do mais recente relatório (AR6 Synthesis Report: Climate Change 2023) produzido pelo Painel Intergovernamental sobre Mudança do Clima (na sigla em inglês, IPCC): as alterações climáticas são uma ameaça ao bem-estar humano e à saúde do planeta (confiança muito elevada). Há uma janela de oportunidade que está se fechando rapidamente para garantir um futuro habitável e sustentável para todos (confiança muito elevada). As escolhas e ações implementadas nesta década terão impactos agora e durante milhares de anos (confiança elevada). (HEADLINE STATEMENTS, 2023).

Como atores centrais do capitalismo contemporâneo, as plataformas são responsáveis por um volume crescente de emissões de dióxido de carbono (Co₂), por um alto consumo de água e por outras ocorrências que estão intrinsecamente vinculadas às suas infraestruturas orientadas por dados (Vatanparast, 2020). O caso mais emblemático é o dos *data centers* (Hogan, 2015; Meyer, 2023): responsáveis pela hospedagem de grandes volumes de dados “na nuvem” - metáfora estrategicamente usada para atribuir leveza a uma atividade fortemente material -, estas infraestruturas consomem cerca de 2% da eletricidade global. A projeção é que este percentual seja quatro vezes maior em 2030. Estima-se que o “setor de tecnologia” tenha sido responsável por 2 a 3% das emissões de carbono em escala mundial em 2021 - percentual equivalente ao do setor de aviação (Navarro, 2023).

No caso das ‘Big Five’ (Alphabet/Google, Amazon, Apple, Microsoft e Meta), o consumo global de eletricidade é equivalente ao da população da Nova Zelândia. Dentre estas plataformas infraestruturais, a recordista em emissões de dióxido de carbono é a Amazon: a expansão de seus negócios durante a pandemia de Covid-19 resultou em um aumento de 18% de emissões entre 2020 e 2021 (Calma, 2022). Não foram localizados dados relativos ao Brasil.

Mas as plataformas não são “apenas” responsáveis por um crescente volume de emissões e consumo de recursos naturais: estão diretamente envolvidas em diversos projetos e ações de sustentabilidade, por exemplo colaborando com a elaboração e aperfeiçoamento de modelos estatísticos baseados em grandes volumes de dados que visam mensurar a “pegada de carbono” de empresas e de atividades econômicas. Através de sistemas de Inteligência Artificial (IA), procuram contribuir para a otimização de estudos e de monitoramentos feitos por instituições ligadas ao meio ambiente, e também dedicam-se a previsões e simulações das alterações socioambientais no curto, médio e longo prazos. Deste modo, as plataformas buscam se posicionar como atores fundamentais nos processos de financiamento e de produção de conhecimento voltados para mitigação ou mesmo solução de desafios da emergência climática.

Esta relação paradoxal das plataformas - pretendem ser parte ativa das ‘soluções’ ao mesmo tempo em que são, de forma cada vez mais intensa, parte do problema - tem sido tematizada por um conjunto ainda pequeno número de estudos. Dwivedi *et al* (2022), por exemplo, mencionam o “paradoxo de Jevons” para apontar que à medida que as melhorias tecnológicas aumentam a eficiência com a qual um recurso é usado, o consumo total desse recurso pode aumentar em vez de diminuir. Já Dobbe e Whitaker (2019), ao repercutirem a adesão de trabalhadores do setor de tecnologia à Greve Global pelo Clima (Tech Workers Coalition, 2019), destacam, entre outros aspectos, o impacto desproporcional com que a instalação de *data centers* atinge comunidades marginalizadas e as parcerias das *Big Techs* com empresas do setor de combustíveis fósseis.

Estes exemplos nos ajudam a situar a crescente vinculação entre plataformas infraestruturais e a crise climática em curso, fenômeno aqui denominado como “plataformização da emergência climática”. Para investigar como as lógicas computacionais, econômicas e ideológicas de *Big Techs* agem tanto para agravar quanto para propor soluções para os desafios socioambientais contemporâneos, neste artigo realizamos um estudo exploratório das ações de “sustentabilidade” divulgadas pelo blog da Google Brasil em 2020 e 2023. Reconhecidamente um “ponto obrigatório de passagem” (Callon, 1986) da internet em função de seus serviços (buscas, email, mapas etc) e de suas infraestruturas (o sistema operacional Android, Google Classroom e Google Cloud, entre outros), a plataforma infraestrutural vinculada à Alphabet² afirma, nos

² Alphabet é o nome de um conglomerado de empresas criado em 2015 com objetivo

3 A neutralidade de carbono está associada à capacidade de compensar a quantidade de CO₂ produzida, enquanto emissão zero de carbono depende da redução drástica de emissões de todos os os gases de efeito estufa

termos destacados em sua página de Sustentabilidade, estar “acelerando a ação climática”. As ações da Google notabilizam-se pelos discursos institucionais e pelo desenvolvimento de soluções tecnológicas - ou, como propomos no título deste trabalho, tecnosoluções - em defesa da sustentabilidade (Whittaker, 2022), mas também por críticas e denúncias publicizadas principalmente através de relatórios e de reportagens. Enquanto a empresa afirma ser, desde 2007, “a primeira grande empresa carbono neutra” e que até 2030 será “a primeira empresa com todos os seus negócios livres de carbono em todo o mundo”³ (Florissi, 2023), estudo conduzido pelo *New Climate Institute* and *Carbon Market Watch* apontou níveis “baixo” (2022) e “moderado” (2023) de transparência e integridade na documentação sobre sustentabilidade (The Climate Costs of..., 2023).

A perspectiva analítica adotada neste estudo ancora-se principalmente nos Estudos de Infraestrutura, um subcampo dos Estudos de Ciência e Tecnologia (STS, na sigla em inglês) que argumenta que esses construtos físicos duradouros devem ser compreendidos não como algo estático, mas sim através articulações entre as múltiplas materialidades e práticas a partir das quais se (re)organizam relações cotidianas e atividades econômicas (Bowker et al, 2019; Vailati & D’andrea, 2020). Quando analisadas na perspectiva STS e abordagens afins, as plataformas infraestruturais são vistas como resultados (não estabilizados) de múltiplos esforços de desenvolvimento científico e tecnológico que incorporam e reorganizam relações de poder. Alinhados a essa visada, argumentamos que plataformas infraestruturais como a Google podem ser vistas são parte da infraestrutura de conhecimento que busca compreender e agir sobre as questões climáticas (Edwards, 2010).

Plataformas infraestruturais e emergência climática

Uma das características centrais das plataformas online é sua intrínseca vinculação com os processos de produção, processamento e intercâmbio de dados (Helmond, 2019|2015). Igualmente importante é o reconhecimento de que sua atuação em escala global não “apenas” depende de robustas infraestruturas como cabos submarinos e servidores de “nuvem”: de modo crescente, as plataformas elas também operam como

infraestruturas a partir das quais são viabilizadas e geridas múltiplas relações comerciais e sociais. Este fenômeno é chamado por Plantin et al. (2018) de “infraestruturização das plataformas”.

Esta configuração é especialmente importante no caso de um seleto grupo de empresas do setor de tecnologia que, em função da forte expansão de atividades e da complexificação de suas bases sociomateriais, viabilizam serviços online de outras empresas e permitem a articulação de variados públicos (como usuários finais, desenvolvedores, produtores de conteúdo e anunciantes). A Amazon (sobretudo através de seu serviço de “nuvem”, a Amazon Web Service), a Meta (através da gestão integrada de aplicações como Facebook, Instagram e WhatsApp e da aposta no “metaverso”) e a Alphabet/Google (Google Cloud e sistema operacional Android, para citarmos apenas dois exemplos) são exemplos de *Big Techs* que, em estudo anterior (d’Andréa, 2023), caracterizamos como plataformas infraestruturais.

Atentas/os a este cenário, um crescente número de pesquisadoras/os de diferentes campos do conhecimento - entre os quais comunicação e mídia (Parks e Starosielski, 2015) - vêem, nos últimos anos, ampliando os diálogos com os Estudos de Infraestrutura, uma frente de estudos ligadas ao campo STS que, desde os anos 1990, dedica-se a compreensão dessas materialidades como “algo que emerge para as pessoas nas práticas, conectada a atividades e estruturas” (Star e Ruhleder, 1995, p. 4). Conforme discutem Vailati & D’Andrea (2020), o olhar antropológico para as infraestruturas dá a ver “práticas, imaginários e efeitos de desigualdade” a partir dos quais se constituem as relações de poder e de dominação.

Buscando uma aproximação mais efetiva entre os estudos de Plataformas/Infraestruturas e as questões climáticas, recorreremos ao livro “A vast machine: computer models, climate data, and the politics of global warming”, no qual Paul Edwards (2010) investigou como se constituiu, a partir dos anos 1960, uma “infraestrutura de conhecimento” de escala global voltada para os estudos do clima. Segundo o autor, atividades como coleta de dados, modelagem, testes e outras operações tecnocientíficas permitiram “uma compreensão amplamente compartilhada do clima e das mudanças climáticas” (p.8) que subsidiaram inclusive os estudos sobre o aquecimento global.

Alinhados a essa perspectiva, argumentamos que as plataformas infraestruturais devem ser vistas também como parte da versão contemporânea da infraestrutura de conhecimento em torno das questões climáticas. Atuando como prestadoras de ser-

viços e como desenvolvedoras de aplicações voltadas para diminuição de impactos ambientais gerados por diferentes setores - inclusive o de tecnologia, operam como “infraestruturas epistêmicas” que privilegiam certos modos de conhecimento em detrimento de outros (Munn, 2022).

O desenvolvimento de conhecimento científico e de soluções tecnológicas pelas plataformas infraestruturais está diretamente associada à sua capacidade de estabelecer parcerias com diferentes tipos de instituições. Um exemplo emblemático é a parceria global com a Google anunciada em 2018 pelo Programa das Nações Unidas para o Meio Ambiente. Da combinação entre ciência ambiental, big data e computação em nuvem, surge a promessa de “mudar o modo como vemos nosso planeta” (UN Environment and Google, 2018).

Ao discutir a temática da sustentabilidade são tratadas nas atuais pesquisas de inteligência artificial, Bergmann & Solomun (2021) chamam atenção para como os serviços infraestruturais ofertados pelas *Big Tech* são fundamentais para a efetivação das práticas e discursos de “greenwashing”, ou seja, de apropriação das questões e causas socioambientais para melhorar a imagem de organizações. Os autores destacam a ênfase dada à quantificação dos problemas e no desenvolvimento de “soluções de medição, monitoramento e otimização contínuos voltados para eficiência”. Ao privilegiar um ponto de vista amplo que viabiliza o cumprimento de metas e legitima os discursos corporativos, argumentam os autores, essas ações negligenciam as especificidades das experiências de pessoas e dos lugares mais vulneráveis, evitando assim lidar com as urgências apontadas por defensores de uma “justiça climática”.

Uma crítica semelhante é feita por Brodie (2023): para discutir a associação entre as infraestruturas de dados e o capitalismo contemporâneo, o autor destaca o papel dos data centers como “infraestruturas materiais e discursivas” que articulam as “transições” de organizações tanto para uma economia baseada em dados e quanto para o uso de “energia verde”. Assim, argumenta Brodie (2023, p.9), “os data centers (...) são simultaneamente produtos energéticos e as infraestruturas das transições energéticas”, o que amplia a centralidade das *Big Techs* não apenas nas ações e investimentos, mais também nas disputas em torno dos “imaginários energéticos” (Camelo, 2018) que orietam decisões políticas.

Esse excessivo foco na criação e no aperfeiçoamento de métricas, aplicações, métodos é lida por críticos como mais uma faceta do tecnosolucionismo que caracteriza a postura das Big Techs frente a diferentes problemas contemporâneos (Morozov, 2018). O tecnosolucionismo ambiental singulariza-se por uma forte aproximação com a economia verde (Nobrega e Varon, 2020), dando suporte a um conjunto mais amplo de medidas que buscam “consertar” problemas relativos à crise climática ignorando as contradições estruturais do regime neoliberal (Levidow, 2023).

Uma das contradições deste *modus operandi* é, como mencionamos na introdução deste trabalho, o fato das atividades das plataformas infraestruturais promoverem cada vez mais consumo energético, água e de minerais. Para operarem na temperatura ideal, os data centers consomem grandes volumes de água, ocasionando problemas hídricos nas localidades onde são construídos (Dobbe e Whittaker, 2019; Hogan, 2015; Viana, 2023). O alto consumo de minerais, especialmente em países periféricos como a Bolívia e o Congo (exploração de lítio), é outro fator de degradação socioambiental associado ao setor (Crawford, 2023).

Dentre as várias aplicações que dependem das infraestruturas plataformizadas, chamam a atenção na atualidade o custo energético do treinamento de grandes modelos de aprendizagem de máquina e de outras formas de inteligência artificial. Em sua crítica ao modelo de produção científica que se preocupa sobretudo com o tamanho dos grandes modelos de linguagem (em inglês, LLMs) em desenvolvimento na atualidade, Bender *et al* (2021) enfatizam o crescente custo socioambiental dessa competição empresarial, por exemplo mencionando um estudo da OpenAI (responsável pelas diferentes versões do LLM nomeado de GPT) que aponta que a capacidade computacional demandada para treinar os maiores modelos de aprendizagem profunda aumentou 300,000 vezes entre 2012 e 2018. Neste debate, mesmo tecnologias aparentemente irrelevantes para a questão climática são pertinentes: como discutem Jancovi e Keilbach (2023), as atuais configurações de algoritmos de compressão de vídeo desenvolvidos e/ou adotados por plataformas de *streaming* concentra o consumo de energia nos usuários finais, desonerando os *data centers* e os relatórios de sustentabilidade de empresas como a Netflix.

Outra contradição fundamental neste debate é forte associação das plataformas infraestruturais com os maiores “vilões” da emergência climática: as empresas do setor

1 Representação 2D na qual cada pixel corresponde distribuição de probabilidade do quão próximo encontram-se os pares de resíduos no mapa de distâncias (SENIOR, EVANS, et al., 2020).

de petróleo e outros combustíveis fósseis. Como descreve Zero Cool (2019), a relação entre as Big Tech e as empresas que compõe o grupo Big Oil é de mútua dependência: para as empresas de tecnologia, a contratação de serviço de nuvem pelo setor petrolífero é um aspecto-chave na competição pelo mercado de “nuvem”, enquanto empresas como Chevron e ExxonMobil dependem das aplicações de inteligência artificial e aprendizagem de máquina para aperfeiçoar processos estratégicos - como análise de grande quantidade de dados sobre exploração de poços de petróleo - em seu também competitivo mercado.

4 A relação atualizada de postagens pode ser acessada em

<https://blog.google/intl/pt-br/novidades/iniciativas/#sustentabilidade>

Estudo de caso: sustentabilidade pela google brasil

Para discutir empiricamente essas questões, realizamos um levantamento de temas, projetos, tecnologias etc. afins à sustentabilidade publicizados pela Google Brasil. Identificamos no “Blog da Google Brasil” dez textos vinculados à categoria “Sustentabilidade” - estes foram publicados entre setembro de 2020 e julho de 2023 (mês de encerramento desta coleta de dados)⁴. Foram duas postagens por ano de 2020 a 2022 e quatro nos sete primeiros meses de 2023.

Além do aumento no número de postagens em 2023, outro indício de uma maior atenção da Google Brasil ao tema é a mudança no perfil de autoras/es dos textos publicados a partir de 2022. Três das quatro publicações de 2020 e 2021 são traduções de textos assinados pelo CEO da Google e da Alphabet (Pichai, 2000; Pichai, 2021a; Pichai, 2021b), enquanto o primeiro texto direcionado ao contexto brasileiro é assinado pelo “Diretor do Google.org para América Latina” (Mujica, 2020). A partir de junho de 2022, predominam artigos sobre projetos voltados para o contexto brasileiro (como crescente destaque para a Amazônia, como discutimos mais à frente) assinados por uma maior diversidade de autoras/es - cada um dos oito textos é escrito por um/a profissional diferente.

Os cargos ocupados indicam um maior alinhamento com o tema “Sustentabilidade”: por exemplo, “Diretora marketing” e co-líder do comitê interno de sustentabilidade.

de" (Mau, 2022), "Diretor do Google for Startups na América Latina" (Barrence, 2022) e "Gerente de Parcerias da área de impacto social do Google" (Mori, 2023). A última postagem é uma síntese da nova edição de relatório ambiental de alcance global que é produzido desde 2016 (Google's 2023 Environmental Report) (Gomes e Brandt, 2023).

Uma análise qualitativa das postagens aponta uma significativa mudança de postura e de foco da Google ao longo dos quatro anos. Na primeira postagem de "Sustentabilidade", Pichai (2020) discursa sobre a "terceira década de ações climáticas" da empresa anunciando o cumprimento da meta de remover todo o legado de carbono emitido antes de 2007 (ano em que a Google tornou-se neutra na emissão de carbonos). O CEO da Google e da Alphabet assume ainda um novo compromisso público: até 2030, operar "com energia limpa 24 horas por dia, 7 dias por semana" - em outras palavras, tornar-se uma empresa "net zero". Para se tornar a "primeira grande corporação global" a cumprir esta meta, Pichai (2020) apresenta medidas internas em curso, como a adoção de Inteligência Artificial e aprendizagem de máquina voltadas para a redução de consumo da energia gasta nos *data centers*. Pichai (2020) menciona ainda a existência de "várias iniciativas de colaboração" voltadas para a redução do consumo energético de parceiros e organizações, de apoios a uma "rede de entidades ambientais" e a implementação gradativa de "produtos" que permitem indivíduos fazerem "escolhas mais sustentáveis", principalmente no deslocamento urbano.

Algumas destas frentes de atuação são retomadas nos posts seguintes assinados pelo CEO. Em Pichai (2021b) é mencionado o uso de IA para analisar as "condições do tráfego e para melhorar o sistema de semáforos" em Israel - com previsão de ampliação para o Rio de Janeiro e outras cidades. Mais destacado é o anúncio de que seria possível a partir de então "tomar decisões" sobre viagens aéreas baseadas na consulta, através da ferramenta Google Voos, às "emissões de carbono associadas por assento para cada voo". (Pichai, 2021b)⁵. Outra informação agregada a um aplicação é a "navegação para ciclistas" no Google Maps em mais de 300 cidades. A dimensão infraestruturais das medidas adotadas então pela Google evidencia-se ainda mais no anúncio de que clientes do Google Cloud poderiam a partir de então escolher em qual servidor preferem alocar seus dados tendo como critério o volume de emissões de carbono de cada *data center* da empresa (Pichai, 2021a)⁶.

O primeiro edital de financiamento a projetos de sustentabilidade anunciado pela Google no Brasil buscou incentivar o uso de dados públicos para "reduzir as emis-

5 Mudanças na metodologia de cálculo de emissões de carbono em voos é uma das controvérsias tecnoambientais que envolvem a Google. Em 2022, a empresa simplificou o "modelo de impacto de viagens" (código disponível em <https://github.com/google/travel-impact-model>), excluindo variáveis cuja mensuração é menos precisa (como o rastro deixado pelas turbinas dos aviões) e mantendo apenas o cálculo de CO₂ emitido na queima de combustível. O volume de emissões mensurados diminuiu quase pela metade, gerando críticas de pesquisadores e governos (ROWLATT, 2022).

6 O consumo estimado de cada data center da Google é informado na página <https://cloud.google.com/sustainability/reunion-carbon>

7 <https://plataforma.alerta.mapbiomas.org/>

8 <https://plataforma.brasil.mapbiomas.org/pistas-de-pouso>

9 <https://imazon.org.br/categorias/sad/>

10 <https://imazon.org.br/imprensa/entenda-o-sistema-de-monitoramento-da-exploracao-madeira-simex/>

11 <https://sites.research.google/floods/>

12 <https://www.cria.org.br/>

13 <https://oncafari.org/>

14 <https://www.nature.org/en-us/>

sões e melhorar a qualidade do ar nas cidades de Curitiba e Porto Alegre.” (Mujica, 2020). Outra iniciativa de financiamento identificada no período analisado foi o edital “Google for Startups Accelerator” - Sustentabilidade, lançado em 2022. Mentoria para desenvolver um projeto técnico e “acesso a créditos para utilizar produtos do Google focados em Cloud, Machine Learning, Android e Web” são alguns dos benefícios do edital (Mau, 2022). Dez startups foram contempladas no edital, entre as quais um “marketplace de energia renovável” e “uma startup que utiliza dados de satélites e nano satélites para monitorar remotamente ameaças florestais” (BARRANCE, 2022).

Mais do que os editais de financiamento, chama a atenção, especialmente em 2023, as parcerias feitas pela Google com diferentes instituições ligadas ao meio ambiente, o que inclui órgãos governamentais. O apoio infraestrutural é o foco dessas parcerias, com destaque para hospedagem de grandes volumes de dados em *data centers* da Google e para utilização de serviços agregados a eles, como o Google Earth Engine, uma plataforma de análise geoespacial que “combina dados de centenas de satélites e conjuntos de dados de observação da Terra com computação em nuvem”. Nas palavras do presidente do Google no Brasil, “com o poder do Google Cloud e com a inteligência do Earth Engine, MapBiomas e Imazon estão criando informações valiosas e insights que contribuem para os esforços contra o desmatamento” (Coelho, 2023).

Vinculado ao Observatório do Clima e produzido por uma rede colaborativa de instituições, o MapBiomas mantém, com apoio do Google, serviços como o Map Biomas Alerta⁷ (autodefinido como “sistema de validação e refinamento de alertas de desmatamento de vegetação nativa em todos os biomas brasileiros com imagens de alta resolução”) e o Mapa das Pistas⁸, voltado para identificação de pistas de pouso. A parceria foi firmada em 2015 e, nas palavras de Florissi (2023), se configura como “um dos maiores casos de uso globais de nossa plataforma de análise geoespacial (o Google Earth Engine)”.

Já o Instituto do Homem e Meio Ambiente da Amazônia (Imazon) conta o apoio da Google para viabilizar o funcionamento do Sistema de Alerta de Desmatamento (SAD)⁹, uma “ferramenta de monitoramento da Amazônia Legal baseada em imagens de satélites” que acompanha a degradação florestal e do desmatamento na região, e do Sistema de Monitoramento de Exploração Madeireira (Simex), que se concentra no estudo de áreas submetidas à exploração madeireira na Amazônia¹⁰. Além disso, a Google ofe-

rece créditos na nuvem para a colaboração do Instituto de Matemática Pura e Aplicada (IMPA) com o Imazon visando o desenvolvimento de “algoritmos de inteligência artificial para automatizar o processo de análise dos alertas de desmatamento” (Florissi, 2023).

Dentre as parcerias com órgãos governamentais brasileiros estão o apoio ao Instituto Nacional de Pesquisas Espaciais (INPE) para aperfeiçoamento da detecção de incêndios florestais através de imagens de satélite e também ao Serviço Geológico do Brasil (órgão vinculado ao Ministério de Ciência e Tecnologia) através da implementação, no fim de 2022, do programa Alerta de Previsão de Cheias¹¹ de rios (Mori, 2023). Dois outros projetos viabilizados pela infraestrutura Google dedicam-se à identificação de elementos da flora e da fauna brasileiras: pré-classificação de milhões de imagens de plantas armazenadas pelo Centro de Referência em Informação Ambiental (CRIA)¹² e automatização da identificação de onças que vivem em unidades de conservação monitoradas pela Associação Onçafari¹³ (Florissi, 2023).

Por fim, o uso de “aprendizado de máquina e bioquímica para criar uma tecnologia inovadora” é a base de um projeto focado no desenvolvimento de uma API de dados abertos voltada para o rastreamento da origem da madeira da Amazônia comercializada no Brasil e no exterior. Chamada de “Digitais da Floresta”, a aplicação inclui um mapeamento das “assinaturas químicas” de árvores da floresta e posterior uso de modelos preditivos de IA para cruzar esta informação com outros dados visando determinar o local de extração da madeira. O projeto é coordenado pela The Nature Conservancy (TNC)¹⁴ e conta com apoio financeiro e de funcionários da Google (Coelho, 2023).

Considerações finais

A síntese apresentada acima das ações de “sustentabilidade” endossadas pela Google Brasil nos permite fazer alguns apontamentos e sinalizar questões a serem desdobradas em pesquisas futuras. Nas primeiras postagens, destacam-se as ações globais divulgadas pelo CEO da Google e da Alphabet (Pichar, 2020). Estas ilustram especialmente a forte associação entre as *Big Techs* e a transição do setor energético, conforme discute Brodie (2023). O aumento da eficiência no consumo de energia é apresentado

como fortemente vinculado (ou mesmo dependente) de sistemas de Inteligência Artificial.

Ao longo dos três anos analisados, uma mudança chama atenção: se, nas primeiras postagens, a atenção da Google Brasil parece estar centrada em grandes cidades (incluindo algumas brasileiras), a partir de 2023 grande atenção passa a ser dispensada à Amazônia. O foco principal é apoiar organizações especializadas no monitoramento socioambiental através de imagens de satélites, atividade esta que depende do processamento de grandes volumes de dados (viabilizado, no caso, através do Google Cloud) e de aplicações como a plataforma de análise geoespacial Google Earth Engine.

Assim, destacamos como, assim como os setores de energia, de petróleo e diversos outros, também o combate ao desmatamento, as previsões climáticas e outras atividades de monitoramento socioambiental parecem ser cada vez mais dependentes dos serviços infraestruturais providos pelas plataformas. Nesse cenário, cabe ainda observar a existência de uma certa disputa entre as *Big Techs* no que tange à presença na região amazônica. Em maio de 2022, por exemplo, Elon Musk se encontrou com o então presidente brasileiro Jair Bolsonaro para discutir o uso da rede Starlink para prover acesso à internet a escolas em áreas rurais e para realizar monitoramentos ambientais (Elon Musk visits, 2022).

Esta análise inicial da plataformização da emergência climática através das iniciativas da Google torna visível os vínculos entre seus interesses comerciais e as ações de empresas de diferentes setores - daquelas interessadas em otimizar uma "transição energética" às instituições responsáveis pelo monitoramento e controle socioambiental. Posicionando-se como "ponto obrigatório de passagem" na busca global por tecnosolucionismos ambientais (Nobrega e Varon, 2020), as plataformas infraestruturais operam também com imaginários e visões de futuro (Camelo, 2018), e assim contribuem para a consolidação da perspectiva político-ideológica que atribui às inovações tecnológicas a capacidade de resolver - de modo "acelerado", para lembrar o slogan na página de Sustentabilidade da Google - problemas criados pela própria lógica tecnocientífica que as orienta.

Referências

BARRENCE, A. O propósito do Google for Startups em acelerar empresas com soluções sustentáveis. Blog do Google Brasil. Disponível em: <<https://blog.google/intl/pt-br/novidades/iniciativas/o-proposito-do-google-for-startups-em-acelerar-empresas-com-solucoes-sustentaveis/>>. Publicado em: 13 set. 2023. Acesso em: 13 jun. 2023.

BERGMANN, R; SOLOMUN, S. A New AI Lexicon: Sustainability. AI Now Institute, 18 out. 2021. Disponível em: <<https://ainowinstitute.org/publication/a-new-ai-lexicon-sustainability>>. Acesso em: 18 jun. 2023

BOWKER, G. C. et al. Introduction to Thinking Infrastructures. In: KORNBERGER, M. et al. (Eds.). *Thinking Infrastructures. Research in the Sociology of Organizations*. [s.l.] Emerald Publishing Limited, 2019. v. 62p. 1–13.

BRODIE, P. Data infrastructure studies on an unequal planet. *Big Data & Society*, v. 10, n. 1, 1 jan. 2023.

CALLON, M. Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fishermen of St Brieuc Bay. *The Sociological Review*, v. 32, n. 1_suppl, p. 196–233, 1 maio 1984.

CALMA, J. Big Tech is pouring millions into the wrong climate solution at Davos. *The Verge*, 25 mai. 2022.

CAMELO, Ana Paula. Futuros energéticos no Antropoceno: trazendo as dimensões sociais para o debate. *ClimaCom Cultura Científica - pesquisa, jornalismo e arte*. Ano 5 - N 12, Julho de 2018. p.29-46

CHAN, Kelvin. Google cracks down on climate change denial by targeting ads. *Associated Press*. 08 out. 2021. Disponível em <<https://apnews.com/article/climate-change-technology-business-misinformation-environment-and-nature>>. Acesso em: 13 jun. 2023.

COELHO, F. Sustentabilidade com o Google: ajudando a preservar a Amazônia e a construir um futuro mais sustentável. Blog do Google Brasil. Disponível em: <<https://blog.google/intl/pt-br/novidades/iniciativas/sustentabilidade-com-google-ajudando-preservar-a-floresta-amazonica/>>. Publicado em: 04 abr. 2023. Acesso em: 13 jun. 2023.

CRAWFORD, K. *Atlas of Ai: power, politics, and the planetary costs of artificial intelligence*. New Haven: Yale University Press, 2021.

CRISTOFARI, G. *The Politics of Platformization – Amsterdam Dialogues on Platform Theory*. Amsterdam: Institute of Network Cultures, 2023.

DOBBE, Roel, WHITTAKER, Meredith. *AI and Climate Change: How they're connected, and what we can do about it*. AI Now Institute, 17 out. 2019. Disponível em: <<https://ainowinstitute.org/publication/ai-and-climate-change-how-theyre-connected-and-what-we-can-do-about-it>>. Acesso em: 18 jun. 2023

DWIVEDI, Y. K. et al. *Climate change and COP26: Are digital technologies and information management part of the problem or the solution? An editorial reflection and call to action*. *International Journal of Information Management*, v. 63, p. 102456, 1 abr. 2022.

EDWARDS, P. N. *A vast machine: computer models, climate data, and the politics of global warming*. Cambridge, Mass: MIT Press, 2010.

EDWARDS, P. N. *Platforms Are Infrastructures on Fire*. In: MULLANEY, T. S. et al. (org.). *Your Computer Is on Fire*. Boston, The MIT Press, 2021

Elon Musk visits Brazil to discuss Amazon with Bolsonaro. *CNBC*, 20 maio 2022.

FLORISSI, Patricia. *Como a nuvem está ajudando a monitorar e preservar a Amazônia*. Blog da Google Brasil. 04 Abr. 2023. Disponível em <<https://blog.google/intl/pt-br/ produtos/nas-nuvens/como-a-nuvem-esta-ajudando-a-monitorar-e-preservar-a-amazonia/>>

INFLUENCE MAP - *Climate Change and Digital Advertising - The Oil and Gas Industry's Digital Advertising Strategy*. Agosto de 2021. Disponível em <<https://influencemap.org/report/Climate-Change-and-Digital-Advertising-a40c8116160668aa2d865da2f5abe91b#1>>. Acesso em 30 jun. 2022

GOMES, Ben, BRANDT, Kate. *Nosso relatório ambiental 2023*. Blog da Google Brasil. 24 jul. 2023. Disponível em: <<https://blog.google/intl/pt-br/novidades/nosso-relatorio-ambiental-2023/>>. Acesso em: 1 ago. 2023.

HELMOND, A. *Plataformização da Web*. In: OMENA, J. (ed.) *Métodos Digitais: teoria-prática-crítica*, 2015|2019. p. 49-76

HOGAN, M. *Data flows and water woes: The Utah Data Center*. *Big Data & Society*, v. 2, n. 2, p. 2053951715592429, 1 dez. 2015.

JANCOVIC, M.; KEILBACH, J. 4. *Streaming against the Environment: Digital Infrastructures, Video Compression, and the Environmental Footprint of Video Streaming*. Em: ES, K.; VERHOEFF, N. (Eds.). *Situating Data*. [s.l.] Amsterdam University Press, 2023. p. 85-102.

KIRCHGAESSNER, Stephanie. *Revealed: Google made large contributions to climate change deniers*. *The Guardian*. 11 nov. 2019. Disponível em <<https://www.theguardian.com/environment/2019/oct/11/google>>

-contributions-climate-change-deniers >

LEE, M. *Alphabet: The Becoming of Google*. New York: Routledge, 2019.

MAYER, V. When do we go from here? Data center infrastructure labor, jobs, and work in economic development time and temporalities. *New Media & Society*, v. 25, n. 2, p. 307–323, 1 fev. 2023.

MAU, M. Google for Brasil: Renovando nosso compromisso com a sustentabilidade no Brasil. Blog do Google Brasil Disponível em: <<https://blog.google/intl/pt-br/novidades/iniciativas/google-for-brasil-renovando-nosso-compromisso-com-a-sustentabilidade-no-brasil/>>. Publicado em: 14 jun. 2022. Acesso em: 13 jun. 2023.

MUJICA, H. Apoiando cidades brasileiras a gerar impacto positivo sobre o clima. Blog do Google Brasil Disponível em: <<https://blog.google/intl/pt-br/novidades/iniciativas/apoiando-cidades-brasileiras-gerar-impacto-positivo-sobre-o-clima/>>. Publicado em: 23 out. 2020. Acesso em: 13 jun. 2023.

MORI, Bia. Utilizando Inteligência Artificial para prever inundações, incêndios e combater o desmatamento. Blog do Google Brasil. 10 Abr. 2023. Disponível em <<https://blog.google/intl/pt-br/novidades/tecnologia/utilizando-inteligencia-artificial-para-prever-inundacoes-incendios-e-combater-o-desmatamento/>> Acesso em: 13 jun. 2023.

MOROZOV, Evgny. *Big Tech: A ascensão dos dados e a morte da política*. Ubu Editora. São Paulo, 2018.

NAVARRO, R. The Carbon Emissions of Big Tech. *ElectronicsHub*, 13 fev. 2023.

NOBREGA, Camila; VARON, Joana,. Big tech goes green(washing): Feminist lenses to unveil new tools in the master's houses. In: *Global Information Society Watch 2020: Technology, the environment and a sustainable world: Responses from the global South*. [s.l.] Association for Progressive communications (APc) And swedish international development cooperation Agency (sida), 2021.

PARKS, Lisa; STAROSIELSKI, Nicole. *Signal Traffic: Critical Studies of Media Infrastructures*. Chicago: University of Illinois Press, 2015.

PICHAİ, S. Nossa terceira década de ações climáticas: conquistando um futuro livre de carbono. Blog do Google Brasil. Disponível em: <<https://blog.google/intl/pt-br/novidades/iniciativas/nossa-terceira-decada-de-acoas/>>. Publicado em 14 set. 2020. Acesso em: 13 jun. 2023.

PICHAİ, Sundar. Novos avanços rumo à nossa meta de funcionar usando apenas energia sem CO2. Blog

do Google Brasil. 22 Abr, 2021a. Disponível em <<https://blog.google/intl/pt-br/novidades/iniciativas/novos-avancos-rumo-meta-energia-sem-co2/>> Acesso em: 13 jun. 2023.

PICHAU, Sundar. Oferecendo a você opções mais sustentáveis com o Google. Blog do Google Brasil. 06 Out, 2021b. Disponível em <<https://blog.google/intl/pt-br/novidades/iniciativas/oferecendo-voce-opcoes-mais/>> Acesso em: 13 jun. 2023.

PLANTIN, J.-C. et al. Infrastructure studies meet platform studies in the age of Google and Facebook. *New Media & Society*, v. 20, n. 1, p. 293–310, 1 jan. 2018.

POELL, T.; NIEBORG, D.; DIJCK, J. VAN. Plataformização. *Fronteiras - estudos midiáticos*, v. 22, n. 1, p. 2-10–10, 4 abr. 2020.

ROWLATT, Justin. Google 'airbrushes' out emissions from flying, BBC reveals. *BBC News*. 26 set. 2022. Disponível em <<https://www.bbc.com/news/science-environment-62664981>>

TECH WORKERS COALITION - Climate Strike. 2019. Disponível em <https://techworkerscoalition.org/climate-strike/>

THE CLIMATE COSTS OF BIG TECH. AI Now Institute 11 abr. 2023. Disponível em: <<https://ai-nowinstitute.org/spotlight/climate/>>. Acesso em: 18 jun. 2023

UN Environment and Google announce ground-breaking partnership to protect our planet. *UN Environment Programme*. 16 jul 2018. Disponível em <<https://www.unep.org/news-and-stories/press-release/un-environment-and-google-announce-ground-breaking-partnership>> Acesso em 05 set. 2023

VAN DIJCK, J.. *The Culture of Connectivity: A Critical History of Social Media*. Londres: Oxford University Press, 2013.

VATANPARAST, R. Data Waste. *Harvard International Law Journal Frontiers*, 1 jan. 2020.

VIANA, N. ß A briga dos uruguaios contra o Google. *Agência Pública*. Publicado em : 13 ago. 2023. Disponível em <<https://apublica.org/2023/08/a-briga-dos-uruguaios-contr-o-google/>> . Acesso em: 15 ago. 2023

WHITTAKER, M. Google has one of Big Tech's most aggressive sustainability plans. Here's its 3-step playbook for helping the planet. *Yahoo Finance*, 12 set. 2022.

ZERO COOL. Oil is the New Data. *Logic Magazine*, 07 dez. 2019. Disponível em <https://logicmag.io/>

nature/oil-is-the-new-data/. Acesso em: 13 jun. 2023.



■ Giselle Beiguelman, frames do vídeo generativo da série Flora Rebellis, versão 2024.
Imagens geradas com Inteligência Artificial (Style Gans)

■ [Acesse o vídeo clicando na imagem](#)





A EXPANSÃO DA INTELIGÊNCIA ARTIFICIAL E SEU IMPACTO NAS DINÂMICAS SOCIAIS:

Desafios e responsabilidades:

DOI: <https://doi.org/10.35699/2965-6931.2023.47727>

JULIÃO GONÇALVES AMARAL*

RESUMO: A Inteligência Artificial (IA) está impulsionando uma revolução industrial recente, transformando desde tarefas manuais até aquelas que exigem inteligência. Enquanto entramos numa fase entusiasmada com os benefícios da IA, isso é impulsionado pelo acesso acessível a processamento e memória, paradigmas emergentes como redes neurais profundas e vastos dados disponíveis. No entanto, junto com o entusiasmo, surgem preocupações legítimas. As perspectivas sobre o impacto variam, incluindo o papel das plataformas digitais, a transformação do mercado de trabalho e a digitalização dos negócios. A ética na IA se tornou uma necessidade incontornável, não apenas um debate intelectual, abordando a autonomia compartilhada, a privacidade, a segurança cibernética e os impactos socioeconômicos. Diante desses desafios, a colaboração entre campos, governos e organizações é essencial para estabelecer bases éticas sólidas e garantir que a IA beneficie a humanidade de maneira justa.

PALAVRAS-CHAVE: inteligência artificial; ética na IA; transformação tecnológica; desafios e benefícios sociais..

The expansion of artificial intelligence and its impact on social dynamics: Challenges and responsibilities

ABSTRACT: Artificial Intelligence (AI) is driving a recent industrial revolution, transforming tasks from manual to those requiring intelligence. As we enter an enthusiastic phase about the benefits of AI, this is propelled by affordable access to processing and memory, emerging paradigms like deep neural networks, and vast available data. However, alongside the enthusiasm, legitimate concerns arise. Perspectives on impact vary, including the role of digital platforms, labor market transformation, and business digitization. Ethics in AI has become an unavoidable necessity, not merely an intellectual debate, addressing shared autonomy, privacy, cybersecurity, and socio-economic impacts. In the face of these challenges, collaboration across fields, governments, and organizations is crucial to establish robust ethical foundations and ensure that AI benefits humanity fairly.

KEYWORDS: artificial intelligence; ethics in AI; technological transformation; social challenges and benefits.

* Universidade Federal de Minas Gerais.

Introdução

A crescente ascensão da Inteligência Artificial (IA) está imbuída de uma profunda transformação na forma como vivemos e interagimos com o mundo à nossa volta. Esta revolução tecnológica tem sido impulsionada pelo desenvolvimento e avanço de tecnologias avançadas que desencadearam uma série de mudanças significativas em diversos setores da sociedade. A IA deixou de ser apenas uma ferramenta que executa tarefas específicas para se tornar uma força motriz por trás de um novo paradigma industrial, comparável às revoluções que a humanidade testemunhou no passado.

No epicentro dessa transformação está a capacidade da IA de transcender suas aplicações tradicionais, estendendo-se além das tarefas manuais e adentrando o território das tarefas racionais, onde é necessária uma abordagem inteligente e adaptativa para lidar com situações complexas. Essa mudança de paradigma não apenas promete aumentar a eficiência e a produtividade, mas também redefine a própria natureza do trabalho humano, desafiando as concepções tradicionais sobre a interação entre máquinas e seres humanos.

Nesse contexto, é impossível ignorar o crescente entusiasmo em relação aos inúmeros benefícios potenciais que a IA pode proporcionar. A convergência de fatores, como o custo cada vez mais acessível de processamento e memória, o surgimento de paradigmas inovadores, como redes neurais profundas, e a vasta disponibilidade de dados na era digital, têm contribuído para esse clima de otimismo. Contudo, essa onda de entusiasmo também está acompanhada por uma série de preocupações, muitas delas validamente fundamentadas.

À medida que a IA se estabelece como um pilar central da sociedade moderna, surgem debates intensos e perspectivas diversas sobre o impacto abrangente que ela pode ter. Algumas análises enfatizam o papel das plataformas digitais como agentes de disrupção, pavimentando o caminho para um novo paradigma de capitalismo baseado em plataformas. Por outro lado, outros destacam a fascinante natureza da IA e preveem transformações profundas que podem espelhar momentos históricos de grande mudança, como a Grande Transformação e a Era das Máquinas.

Nesse cenário multifacetado, a complexidade e interconexão das tecnologias emer-

1 Algumas conclusões do mais recente relatório (AR6 Synthesis Report: Climate Change 2023) produzido pelo Painel Intergovernamental sobre Mudança do Clima (na sigla em inglês, IPCC): as alterações climáticas são uma ameaça ao bem-estar humano e à saúde do planeta (confiança muito elevada). Há uma janela de oportunidade que está se fechando rapidamente para garantir um futuro habitável e sustentável para todos (confiança muito elevada). As escolhas e ações implementadas nesta década terão impactos agora e durante milhares de anos (confiança elevada). (HEADLINE STATEMENTS, 2023).

gentes não podem ser subestimadas. A IA não opera em um vácuo, mas sim em conjunto com outras tecnologias-chave, como a computação em nuvem, a Internet das Coisas e o Big Data. Essa interação sistêmica cria não apenas oportunidades sem precedentes, mas também desafios significativos que vão muito além do âmbito puramente tecnológico.

Em última análise, estamos diante de uma encruzilhada crucial em nossa jornada rumo a um futuro moldado pela IA. A busca por um equilíbrio entre os benefícios transformadores e os riscos inerentes é uma tarefa que exige um compromisso coletivo. À medida que a IA se solidifica como uma força impulsionadora, é essencial considerar não apenas os aspectos técnicos, mas também as implicações éticas, sociais e econômicas que moldarão a sociedade que desejamos ver emergir desse processo. Somente através de uma abordagem abrangente e colaborativa poderemos pavimentar um caminho que permita à IA verdadeiramente enriquecer e aprimorar nossa experiência humana.

Este texto tem por objetivo trazer para a discussão alguns dos impactos sociais da IA atualmente. Para além de adotarmos uma posição unicamente otimista ou exclusivamente alarmista, nossa intenção é mostrar os potenciais da IA (positivos e negativos) e, através de algumas discussões no campo da ética e do trabalho, explicitar a complexidade representada pela emergência destas tecnologias. No primeiro tópico, para além desta introdução, buscamos discutir o que é Inteligência Artificial, o que possibilitou a sua atual expansão e algumas de suas aplicações. No segundo tópico, a nossa discussão gira em torno de questões éticas que giram em torno dos usos da IA. No terceiro tópico, abordamos os impactos da IA no mercado de trabalho e na economia global mostrando como a adoção destas tecnologias podem, por um lado, aumentar consideravelmente a produtividade e a geração de riqueza e, por outro lado, levar ao desemprego e ampliar os custos sociais com a realocação das forças de trabalho.

² Alphabet é o nome de um conglomerado de empresas criado em 2015 com objetivo de separar os vários serviços de internet agregados pela plataforma infraestrutural Google das outras atividades comerciais exploradas por seus fundadores e acionistas (LEE, 2019), como a empresa DeepMind, voltada para desenvolvimento de aplicações de Inteligência Artificial.

Inteligência artificial e machine learning: definindo termos

A Inteligência Artificial (IA), que surgiu nos anos 1950, tem sua raiz quase inseparável da origem do computador. O evento considerado como ponto de partida da IA foi a Conferência de Dartmouth College, ocorrida no verão de 1956, e, que estiveram presentes pesquisadores como John McCarthy, Marvin Minsky, Alan Newell e Herbert Simon, considerados pioneiros nesse empolgante campo da Computação (Sichman, 2021). Segundo Neto, Bonacelli e Pacheco (2020), a comunidade de pesquisadores em IA à época se dividiu em dois grupos principais: a abordagem fundamentada em lógica de regras e a abordagem baseada em redes neurais. O primeiro grupo, também conhecido como simbolistas, pré-determinava as regras que o sistema deveria aplicar para solucionar problemas específicos. O segundo grupo adotava uma metodologia de aprendizado, inspirada pela estrutura das redes neurais do cérebro humano. Por meio de exemplos, permitiam que camadas neuronais artificiais determinassem autonomamente o resultado de suas operações. Embora a primeira onda da IA (até os anos 80) tenha causado desapontamento com o aprendizado, a segunda onda, nas últimas décadas, deve muito ao mesmo (Neto, Bonacelli, Pacheco, 2020).

A década de 1990 não foi tão favorável à IA, sendo por vezes chamada de segundo “inverno da IA” devido a controvérsias e redução de financiamento e interesse. No entanto, trouxe avanços conceituais significativos. Embora multifacetados, um eixo importante foi a adoção de formalismos baseados em probabilidades, estatísticas e otimização. A teoria probabilística, antes criticada na literatura de IA, ganhou relevância. Gradualmente, métodos probabilísticos e estatísticos passaram a permear a IA (Cozman, 2021).

Mas, afinal de contas, o que é Inteligência Artificial? Inteligência Artificial (IA) é, de modo geral, a criação de sistemas computacionais para executar tarefas complexas, geralmente associadas à inteligência humana (Neto, Bonacelli, Pacheco, 2020; Sichman, 2021). É um campo multidisciplinar da ciência da computação, focado em resolver problemas usando várias técnicas e modelos. O propósito da IA é desenvolver sistemas que superem limitações humanas ou desafios algorítmicos convencionais. (Neto, Bonacelli, Pacheco, 2020; Sichman, 2021). O campo de IA é uma gama de

modelos e técnicas (busca, raciocínio, decisão, percepção, linguagem natural, aprendizado de máquina) que abordam problemas complexos, usando paradigmas como simbólico, conexionista, evolutivo e probabilístico (Sichman, 2021).

As IA's, basicamente, se dividem em três tipos: IA Focada, IA Generalizada e IA Superinteligente. Na IA Focada, ou IA Fraca, algoritmos resolvem problemas específicos com grande quantidade de dados. Exemplos incluem Sistemas Especialistas e de Recomendação. A IA Generalizada, também chamada de IA Forte, cria algoritmos comparáveis aos humanos em várias tarefas, usando Aprendizado de Máquina, como na Visão Computacional. Na IA Superinteligente, algoritmos superam humanos em quase todas as tarefas. A existência de sistemas assim, desenvolvidos com IA, ainda é incerta (Ludermir, 2021).

Nos anos 2000, houve um notável progresso em IA, embora a sociedade como um todo não percebesse bem as atividades ocorridas nos laboratórios. A pesquisa em representação do conhecimento tornou-se mais teórica, abrangendo ontologias, lógicas e métodos racionais de decisão e planejamento para agentes individuais ou múltiplos. Sobretudo, a década foi marcada pelo crescimento contínuo das técnicas chamadas coletivamente de “aprendizado de máquina”. No passado, “aprendizado de máquina” abrangia várias ideias para melhorar sistemas com base em experiências, como feedback do usuário ou análise de bases de dados. A explosão de dados e interesse em soluções práticas geraram um crescimento na mineração de dados. O campo expandiu para aplicações diversas, mantendo fundamentos matemáticos. A ascensão do aprendizado profundo, com redes neurais de muitas camadas, foi marcante após 2012. Redes profundas superaram humanos em tarefas de imagem e texto, incluindo tradução automática (Cozman, 2021).

Podemos elencar três tipos principais de Aprendizado de Máquina: Supervisionado, Não Supervisionado e Por Reforço. No Aprendizado Supervisionado, cada exemplo precisa ter uma resposta desejada (um rótulo). O objetivo é construir um classificador que determine corretamente a classe de novos exemplos. Pode ser classificação (rótulos discretos) ou regressão (valores contínuos). No Aprendizado Não Supervisionado, os exemplos não têm rótulos e o algoritmo os agrupa com base em similaridades de atributos. O significado dos agrupamentos é determinado após análise. No Aprendizado por Reforço, o algoritmo recebe sinais de recompensa ou punição

para respostas corretas ou erradas respectivamente. Baseado em hipóteses, determina se a atividade foi bem-sucedida ou não. É amplamente usado em jogos e robótica. As Redes Neurais Artificiais (RNA), são uma das técnicas de Aprendizado de Máquina que imitam estruturas neurais biológicas e se encontram entre as mais bem-sucedidas atualmente. Esses modelos matemáticos processam informação através de neurônios artificiais em diversas camadas que permitem que a máquina simule o aprendizado humano (Ludermir, 2021).

Atualmente, podemos dizer que a IA cresceu devido a avanços tecnológicos em quatro eixos, a saber: extração, armazenamento, transmissão e processamento de dados. Técnicas de IA atuais requerem alto poder computacional e dados abundantes para resolver desafios complexos. Dados são amplamente disponibilizados por empresas, pessoas e internet das coisas (IoT) e circulam de forma intensa na rede mundial de computadores (Carvalho, 2021; Ludermir, 2021). No primeiro eixo, avanços na extração de dados resultaram em sensores mais rápidos, acessíveis e avançados, como câmeras inovadoras. O segundo eixo, armazenamento, beneficiou-se de materiais novos, expandindo capacidade, confiabilidade e reduzindo custos. No terceiro, a transmissão de dados evoluiu com redução de custos e ampliação da capacidade de meios físicos e impacto das redes de computadores e internet, conectando mais “coisas” do que pessoas. A internet, que conectava indivíduos globalmente, cedeu espaço à “internet das coisas”. No quarto eixo, computadores mais ágeis e acessíveis processam dados em menor espaço e tempo (Carvalho, 2021).

Na última década, a economia baseada em dados consolidou-se com a ascensão dos sistemas tecnológicos dando destaque para três blocos de tecnologia: computação em nuvem, big data e a IA propriamente dita, formando a base do cenário digital atual. A explosão de dados impulsionou um avanço importante no campo da IA. Os avanços no aprendizado de máquina ampliaram a relevância dos dados, estimulando o crescimento das redes de transmissão e armazenamento, e fomentando a computação em nuvem. Fica evidente uma interconexão tecnológica notável entre Big Data, algoritmos de aprendizado de máquina e computação em nuvem (Neto, Bonacelli, Pacheco, 2020).

Temos que o desenvolvimento de processadores aptos para aplicações de IA é crucial, mas igualmente essencial é a distribuição desse poder computacional, fenômeno parecido à disseminação da energia elétrica na terceira revolução tecnológica do século

XX. A computação em nuvem, então, assume importância dado que aplicativos de IA demandam potência e acesso a enormes conjuntos de dados, tornando a nuvem ideal para sua execução. A computação em nuvem envolve mover dados de indivíduos e departamentos para centros de dados remotos, operados por empresas que cobram pelo armazenamento e uso. Usuários individuais podem pagar para armazenar dados em Datacenters complexos, enquanto as empresas de tecnologia oferecem acesso a serviços remotos e a aplicativos. Externalizar não apenas o armazenamento e o poder computacional, mas setores inteiros como legal ou vendas, é uma novidade (Neto, Bonacelli, Pacheco, 2020).

O Big Data é o terceiro elemento importante para compreendermos o atual sucesso e expansão da IA atualmente. Os progressos tecnológicos atuais permitiram que uma quantidade sem precedentes de dados passassem a circular pela rede resultando na formação das técnicas de Big Data, inicialmente caracterizada pelos três “V”:

Volume, Variedade e Velocidade. Volume refere-se à quantidade abundante de dados gerados, variedade abrange os diferentes formatos – imagens, sons, vídeos – e fontes – como, por exemplo, sensores sofisticados para monitorar atividades físicas – destes dados e velocidade diz da rapidez com que esses dados são produzidos, circulam e, em alguns casos, desaparecem na rede (Carvalho, 2021). Por meio da Análise de Big Data, os dados se tornaram fonte de insights e serviços valiosos e cerca de 95% dos dados com potencial de valor são semi-estruturados ou não estruturados (encontram-se em estado bruto sem algum tipo de sistematização que possibilite a análise), como áudio, vídeo e texto (Neto, Bonacelli, Pacheco, 2020). A demanda por ferramentas computacionais cresceu com o advento do Big Data, impulsionando a exploração de dados para obter novo conhecimento valioso, empregado na tomada de decisões (Carvalho, 2021).

Esse cenário facilitou a expansão da IA, saindo dos laboratórios de pesquisa para serem empregados em produtos e serviços, gerando benefícios econômicos e sociais. Empresas criaram produtos mais seguros, acessíveis e personalizados. Órgãos públicos aprimoraram serviços à população e organizações não governamentais passam a usá-la para monitorar ações prejudiciais à sociedade com maior eficácia (Carvalho, 2021).

Um exemplo do uso de IA para a melhoria da eficiência na prestação de serviços é a aplicação crescente na área de diagnósticos automáticos, com resultados pre-

cisos, superando às vezes diagnósticos humanos. A iFlytek desenvolveu um “Médico Assistente Robô” que auxilia no diagnóstico sem substituir médicos. Outra aplicação bem-sucedida da IA é a melhoria contínua de tradutores automáticos, como os fornecidos pelo Google. A tradução de texto é bem-sucedida devido ao grande volume de dados disponíveis. A área de Visão Computacional também avançou consideravelmente. Em 2012, uma Rede Neural alcançou alto desempenho na competição de reconhecimento de imagens ImageNet Challenge e em 2016, sistemas de IA superaram o desempenho humano neste campo. Exemplos no uso desta tecnologia incluem o DeepFace do Facebook, identificando rostos independentemente do ângulo (Ludermir, 2021).

Uma aplicação que une diversas tecnologias são os Sistemas Tutores Inteligentes (STI), que visam ao ensino personalizado. A IA tem sucesso em abordagens individuais, porém ainda não se destaca no ensino colaborativo. Na educação, várias tecnologias de IA são empregadas, incluindo o Processamento de Linguagem Natural (PLN) para tradução, análise e interpretação de textos e voz. Esses sistemas podem ler e resumir conteúdos, criar smartbooks personalizados e oferecer tradução de voz em tempo real. As aplicações de PLN também se conectam ao reconhecimento de emoções (Vicari, 2021).

No campo da justiça, há o emprego de modelos para prever resultados judiciais nos Tribunais de Justiça. Combinando ferramentas de Processamento de Linguagem Natural (PLN) para extrair características dos textos, esses modelos selecionam as informações principais que podem levar a previsões válidas, inserem os dados pré-processados em estruturas de aprendizado de máquina e, por fim, avaliam seus respectivos resultados em comparação com os desfechos judiciais reais do tribunal (Bertalan, Ruiz).

Em resumo, a Inteligência Artificial (IA) emergiu como uma revolução tecnológica multifacetada que transcende as fronteiras da ciência da computação e permeia todos os aspectos da sociedade moderna. A definição de IA evoluiu ao longo do tempo, abrangendo desde sistemas que resolvem tarefas complexas até a imitação de inteligência humana em diversas áreas. A divisão das IA's em diferentes tipos – Focada, Generalizada e Superinteligente – evidencia sua capacidade de evoluir e se adaptar às necessidades em constante transformação. A ascensão do aprendizado de máquina e o papel crucial desempenhado pela computação em nuvem e pelo Big Data abriram caminho para a disseminação e aplicação prática da IA, impulsionando avanços em

diagnósticos médicos, tradução, visão computacional, educação e até mesmo no sistema judicial. No entanto, essa revolução tecnológica também traz consigo desafios éticos e sociais complexos, desde a necessidade de transparência e equidade nas decisões algorítmicas até a preocupação com a privacidade e a regulamentação. À medida que a IA continua a transformar nossa realidade, a colaboração entre cientistas, reguladores, empresas e a sociedade em geral se torna fundamental para garantir um desenvolvimento ético, responsável e benéfico para a humanidade. A busca por um equilíbrio entre os benefícios e desafios da IA é um imperativo para moldar um futuro onde a tecnologia sirva verdadeiramente ao bem-estar e progresso da sociedade global. Os desafios éticos em torno da IA serão abordados no próximo tópico.

Desafios éticos no uso da ia

A inteligência artificial (IA) atingiu um patamar sem precedentes em sua evolução tecnológica, moldando de maneira profunda e constante a vida das pessoas em múltiplos aspectos do cotidiano. Essa inegável influência desperta uma gama complexa de reações, que variam entre apreensões e otimismo, encontrando eco não apenas entre pesquisadores e cientistas, mas também entre políticos, ativistas, filósofos, artistas e na população em geral. O impacto da IA transcende barreiras disciplinares, impulsionando discussões que permeiam questões cruciais sobre o futuro da sociedade e a coexistência entre humanos e máquinas.

Nesse cenário, os debates acerca dos princípios éticos que cercam o desenvolvimento e uso da IA assumem uma relevância ímpar. A crescente integração da IA em nossas vidas levanta uma série de interrogações sobre as implicações éticas associadas à automação de processos complexos, tomada de decisões autônomas e interações entre sistemas inteligentes e seres humanos. Essa reflexão ética ganha centralidade não apenas como uma precaução diante de dilemas emergentes, mas como uma oportunidade para orientar a evolução da IA de maneira alinhada com valores humanos e sociais, assegurando que as tecnologias sejam projetadas e utilizadas de forma a beneficiar a humanidade como um todo.

Dentre os elementos éticos que orbitam a IA, destaca-se a necessidade de garantir transparência e responsabilidade nas ações dos sistemas automatizados. A opacidade

dos algoritmos de IA pode levantar preocupações sobre a tomada de decisões não compreensíveis ou justificáveis, afetando a confiança nas tecnologias e gerando riscos de discriminação e viés algorítmico. A questão da equidade também emerge como um ponto crucial, já que a IA pode perpetuar ou até mesmo agravar desigualdades existentes se não forem implementados mecanismos de correção e supervisão adequados. A busca por assegurar que os benefícios da IA sejam distribuídos de maneira justa e acessível torna-se uma missão ética urgente, que exige um esforço conjunto da sociedade, governos e empresas para estabelecer diretrizes e regulamentações que promovam a equidade e evitem danos sociais. Neste tópico, iremos discutir alguns elementos éticos que giram em torno da IA atualmente.

A IA pode causar efeitos positivos e negativos. Pode, por exemplo, evitar tarefas perigosas para humanos, possibilitando a execução por máquinas. Pode também liberar humanos de tarefas automáticas, permitindo focar em atividades mais desafiadoras e gratificantes. A aplicação da IA traz benefícios diversos: melhorias na saúde; avanços na educação; energia acessível e sustentável; detecção de fraudes; transporte mais seguro (apps de transporte), rápido (rotas otimizadas) e sustentável (Ludermir, 2021).

Contudo, a IA também traz impactos negativos, incluindo perda de empregos, agravando desigualdades sociais fazendo com que planejamento e ação governamental sejam necessários para minimizar esses efeitos. Além de preocupações sociais, questões éticas surgem com o uso da IA, como armas automáticas, invasão de privacidade, falta de transparência no uso de informações e decisões não explicáveis (Ludermir, 2021). Entender as complexas redes neurais, por exemplo, é bastante difícil e a opacidade dos algoritmos tem impacto na interpretabilidade de sistemas de decisões práticas (Cozman, 2021).

Em meio a todo o alvoroço causado pela IA, um tema ganha grande relevância no debate: a regulação na pesquisa, produção e uso da IA. Regular a IA é um desafio complexo. Não é apenas uma questão tecnológica, pronta para se adaptar às novas criações. Também é uma questão social, necessitando da expertise de especialistas em ciências humanas para considerar suas várias ramificações. Uma regulação eficaz exige debates sociais, consideração de diferentes perspectivas e clareza duradoura (Carvalho, 2021).

Nesse contexto, organizações como a Unesco e a Unicef têm destacado a importância de educar sobre o uso responsável da IA, especialmente para crianças, adolescentes e idosos, grupos mais vulneráveis a possíveis aplicações mal-intencionadas da tecnologia. No Brasil, leis como a Lei Geral de Proteção de Dados Pessoais (Lei n. 13.709 de 2018) e o Marco Civil da Internet (Lei n. 12.965 de 2014) tratam parcialmente dessas questões, mas não abordam totalmente o viés dos sistemas ou o seu uso questionável (VICARI, 2021).

Para além da regulação, profissionais estão buscando modelos de Aprendizado de Máquina precisos, justos e explicáveis. Essa área é conhecida como Aprendizado de Máquina Confiável. Alguns pesquisadores propõem explicar modelos complexos com abordagens mais simples, enquanto outras exploram Técnicas de Explicações Definidas a Posteriori, embora ainda não sejam totalmente confiáveis (Ludermir, 2021).

Em sua obra seminal sobre IA Ética, Virginia Dignum (2019) apresenta de maneira cativante como a IA deve ser desenvolvida e empregada de forma responsável. A autora advoga pela adoção de uma abordagem ética em três áreas distintas, no processo de elaboração desses sistemas em que as equipes devem antecipar as possíveis consequências para indivíduos e sociedades, garantindo consideração ética; no design do comportamento desses sistemas, em que se torna fundamental incorporar capacidades de raciocínio ético nos agentes inteligentes; e no código de conduta de projetistas e desenvolvedores, em que regulação e certificação se tornam essenciais para assegurar comportamento adequado, seguindo padrões semelhantes a outras profissões (Sichman, 2021).

Para a primeira dimensão (ética no design), a autora propõe a abordagem ART of AI, que garante a inclusão explícita e sistemática de valores humanos e princípios éticos nos processos de design de forma transparente. Essa abordagem é composta pela responsabilização (accountability), que refere-se à necessidade de os sistemas de IA explicarem e justificarem suas decisões e ações para parceiros, usuários e outros envolvidos; responsabilidade (responsibility), que diz respeito ao papel das pessoas e à capacidade dos sistemas de IA de assumirem decisões e identificarem erros – à medida que a cadeia de responsabilidade cresce, é vital vincular decisões da IA ao uso justo de dados e ações das partes interessadas –; e transparência (transparency), que abrange a descrição, inspeção e reprodução dos mecanismos usados pelas IAs para tomada de decisões e adaptação ao ambiente – os algoritmos atuais são em grande

parte caixas-pretas, mas reguladores e usuários exigem explicação e clareza nos dados usados (Sichman, 2021).

Os sistemas de IA transparente integram procedimentos de explicação que possibilitam aos usuários obter informações sobre os modelos e escolhas do sistema. Um tema central, dada sua atual relevância, é como incorporar explicações às técnicas de aprendizado profundo, frequentemente fundamentadas em modelos de caixa-preta, cujos parâmetros internos e valores são pouco compreendidos pelos usuários (Sichman, 2021). Uma IA transparente possibilita que os modelos provenientes de algoritmos de IA, usados para decisões que impactam a vida das pessoas, sejam facilmente compreendidos. Desse modo, a maneira como o modelo chegou a uma decisão deve ser suficientemente nítida para que qualquer indivíduo possa compreender. Exemplos dessas decisões incluem determinar se alguém é elegível para um empréstimo financeiro, se uma pessoa está enfrentando uma determinada condição de saúde ou se um paciente deve seguir o tratamento A ou B (Carvalho, 2021).

A transparência está fortemente ligada à confiança que as pessoas depositam nas decisões feitas por modelos criados a partir de algoritmos de IA. Quanto maior a transparência de um modelo, maior é a confiança que as pessoas têm em suas decisões. Além disso, a transparência possibilita a validação das decisões tomadas pelos modelos. A importância dessa transparência é reforçada na área de IA conhecida como IA explicável. Um dos objetivos dos pesquisadores nesse campo é entender o conhecimento utilizado por modelos de caixa-preta (Carvalho, 2021).

Segundo Sichman (2021), a preocupação com as finalidades de uso de sistemas de IA também tem sido alvo de discussão nos últimos anos. O workshop de Inteligência Artificial para o Bem Social (AI4G 2019), evento associado à IJCAI 2019, ocorrido em Macau, China, teve o propósito de explorar a maneira como a IA pode contribuir para resolver questões sociais. Por sua vez, o workshop de Agentes de Inteligência Artificial Responsável sediado em Montreal, Canadá, reuniu pesquisadores de IA, ética, filosofia, robótica, psicologia, antropologia, ciências cognitivas, direito, estudos de governança regulatória e engenharia para discutir e abordar os desafios complexos relacionados ao design e regulamentação de sistemas de IA. (Sichman, 2021).

A presença da IA nos processos decisórios está cada vez mais ampla. Os algoritmos agora desempenham papéis cruciais em decisões que impactam vidas humanas

e para que as pessoas confiem e aceitem essas decisões, é essencial que percebam sua equidade. São frequentes as notícias sobre decisões tendenciosas e preconceituosas tomadas por modelos de IA. Um exemplo emblemático é o de uma ferramenta de identificação de criminosos baseada em IA que incorporava informações raciais em suas decisões. Como resultado, a ferramenta tendia a identificar indivíduos de certa raça como mais propensos a serem criminosos. Uma das principais fontes de preconceito reside nos conjuntos de dados usados para treinar modelos, como ocorre em aplicações que empregam algoritmos de aprendizado de máquina. Portanto, é crucial ter cautela na coleta desses dados. Quando um algoritmo de aprendizado de máquina é aplicado a um conjunto de dados, ele busca padrões e os reforça ao criar um modelo representativo dos dados. Se esses padrões contiverem viés, o modelo resultante também terá viés (Carvalho, 2021).

A IA equitativa amplia as salvaguardas contra a formação de modelos tendenciosos devido a padrões embutidos em dados pessoais. Ela visa evitar que modelos gerados por algoritmos de IA tomem decisões tendenciosas, considerando aspectos como classe social, religião, condições prévias de saúde, idade, nacionalidade, orientação sexual e raça. O uso de variáveis sensíveis representando esses aspectos pode contribuir para a criação de modelos tendenciosos. Portanto, frequentemente, os valores dessas variáveis são tratados de maneira criptografada (Carvalho, 2021).

Outro ponto crucial para o uso ético da IA é o respeito à privacidade e salvaguarda dos dados. O direito à privacidade dos dados pessoais consiste essencialmente no controle que um indivíduo possui sobre a coleta, uso e compartilhamento de suas informações pessoais. O aumento substancial na geração de dados trouxe à tona frequentes casos de vazamentos acidentais ou ilícitos de informações pessoais (Carvalho, 2021).

Para lidar com o crescente uso e criação de dados, a União Europeia revisou sua legislação em 2016, aprovando o Regulamento Geral de Proteção de Dados (GDPR, sigla em inglês para General Data Protection Regulation). O GDPR estabelece diretrizes para a coleta, armazenamento e utilização de dados pessoais através de diversas normativas. Uma dessas normas é o direito à explicação. De acordo com esse princípio, organizações tanto públicas quanto privadas que operam com dados pessoais de cidadãos da União Europeia são obrigadas a assegurar um processamento justo e transparente desses dados. O GDPR também outorga às autoridades de supervisão europeias

o poder de investigar e impor penalidades. Inspirado no GDPR, o Congresso Nacional brasileiro aprovou em 2018 a Lei Geral de Proteção de Dados Pessoais (LGPD), que entrou em vigor em 2020. Leis como o GDPR e a LGPD contribuem para ampliar a transparência da IA. A fim de fiscalizar o cumprimento da LGPD, foi instituída a Autoridade Nacional de Proteção de Dados (ANPD) em 2018 (Carvalho, 2021).

Muitos são os desafios colocados pela atual expansão da IA. Em um artigo de notável interesse, Thomas Dietterich e Eric Horvitz (2015) identificaram cinco categorias de desafios relacionados ao uso de sistemas de IA. A primeira categoria diz das imperfeições (defeitos) dos sistemas. Todos os sistemas de software estão sujeitos a imperfeições. Diversos deles passaram por desenvolvimento e validação criteriosos para garantir alta qualidade; por exemplo, sistemas de piloto automático e controle de naves espaciais passam por testes e validações minuciosos. Abordagens similares devem ser aplicadas aos sistemas de IA. A segunda diz da segurança cibernética. Sistemas de IA enfrentam as mesmas vulnerabilidades que qualquer outro software em relação a ataques cibernéticos. Por exemplo, ao manipular dados de treinamento ou preferências e trade-offs embutidos em modelos de utilidade, adversários podem modificar o comportamento desses sistemas. A terceira categoria diz, segundo Dietterich e Horvitz (2015), de sistemas “aprendiz de mago” (aprendiz de feiticeiro). Um aspecto crítico de qualquer sistema de IA que interage com seres humanos é sua capacidade de interpretar as intenções das pessoas, em vez de seguir comandos literalmente. Um sistema de IA deve analisar e entender se o comportamento solicitado por um ser humano pode ser considerado “normal” ou “razoável” pela maioria e não criminoso. A quarta categoria diz da autonomia compartilhada. A construção desses sistemas colaborativos levanta um conjunto de desafios relacionados à fluidez da colaboração e à clareza sobre estados internos e objetivos dos envolvidos no sistema. Criar sistemas em tempo real nos quais o controle precise alternar rapidamente entre humanos e sistemas de IA é uma tarefa complexa. A quinta e última categoria de desafios diz dos impactos sócio-econômicos da IA. É crucial compreender as implicações da IA na distribuição de empregos e na economia em um contexto mais amplo. Essas questões ultrapassam os limites da ciência da computação e engenharia, adentrando o âmbito das políticas e estratégias econômicas que podem garantir que os ganhos de produtividade baseados em IA sejam equitativamente compartilhados (Sichman, 2021).

Em um mundo em constante evolução impulsionado pela inteligência artificial (IA), emerge a clara percepção de que a intersecção entre tecnologia e ética é um ponto central para orientar esse avanço e assegurar um futuro harmônico para a sociedade. A IA atingiu um patamar de influência sem precedentes, afetando os mais variados aspectos do cotidiano das pessoas, e essa realidade suscita reações complexas que variam entre esperanças e apreensões. Não é apenas um debate entre especialistas e acadêmicos, mas uma discussão que transcende barreiras disciplinares, reunindo políticos, ativistas, artistas e a população em geral.

No cerne desses debates encontra-se a questão ética que permeia o desenvolvimento e a utilização da IA. Essa reflexão ética não apenas age como um guarda-chuva protetor para prevenir dilemas emergentes, mas também oferece uma oportunidade única de orientar o curso da IA em sintonia com os valores humanos e a justiça social. A necessidade de transparência e responsabilidade nas ações dos sistemas automatizados torna-se premente para mitigar riscos de discriminação e vieses algorítmicos, enquanto a busca pela equidade na distribuição dos benefícios da IA destaca a urgência de diretrizes e regulamentações que garantam benefícios acessíveis e justos.

A complexidade ética que envolve a IA não se limita apenas a seu desenvolvimento, mas se estende a aspectos como privacidade, segurança cibernética e impactos socioeconômicos. A exploração de modelos de Aprendizado de Máquina confiáveis e explicáveis busca garantir a interpretabilidade das decisões tomadas por sistemas de IA, fundamentando a confiança e validação desses sistemas. A legislação, como o Regulamento Geral de Proteção de Dados da União Europeia e a Lei Geral de Proteção de Dados Pessoais do Brasil, destaca a crescente importância da transparência e privacidade no uso da IA.

À medida que a IA se infiltra nas tomadas de decisão, torna-se necessário enfrentar desafios como a autonomia compartilhada e a compreensão dos impactos socioeconômicos da IA. Estes desafios transcendem fronteiras acadêmicas e tecnológicas, demandando colaboração interdisciplinar e uma abordagem global. Assim como alertaram Dietterich e Horvitz (2015), é essencial entender que a IA não é apenas uma ferramenta técnica, mas uma força poderosa que molda o presente e o futuro da sociedade.

Portanto, o diálogo em torno da ética na IA não é apenas um debate intelectual, mas uma necessidade incontornável. À medida que a IA continua a expandir suas fronteiras e influência, a sociedade enfrenta a tarefa imperativa de estabelecer bases éticas sólidas para garantir que os avanços tecnológicos sejam benéficos, justos e alinhados com os valores humanos fundamentais. A colaboração entre campos, governos e organizações é essencial para definir um caminho ético claro e sustentável para a IA, assegurando que suas realizações continuem a elevar a humanidade a novos patamares de progresso e bem-estar.

Os desafios no campo do trabalho e da economia são vários e merecem atenção mais detida ao tema. No próximo tópico, discutiremos mais detidamente sobre os impactos da IA na economia global e no mercado de trabalho.

Os impactos da ia no mercado de trabalho: sonho ou pesadelo?

A inserção do uso da inteligência artificial (IA) no mercado de trabalho tem gerado impactos positivos significativos em diversas áreas. Uma das principais vantagens é o aumento da eficiência e produtividade, visto que a IA é capaz de automatizar tarefas repetitivas e rotineiras, permitindo que os trabalhadores se concentrem em atividades mais criativas e estratégicas. Além disso, a IA pode auxiliar na tomada de decisões mais informadas e precisas, proporcionando insights valiosos a partir da análise de grandes volumes de dados, o que pode contribuir para a melhoria dos processos empresariais e o desenvolvimento de produtos e serviços inovadores. Outro impacto positivo é o potencial de criação de novas oportunidades de emprego, uma vez que a expansão da IA demanda profissionais especializados em sua implementação, manutenção e otimização, impulsionando o crescimento de setores como engenharia de dados, desenvolvimento de algoritmos e análise de IA.

Por outro lado, a adoção da IA também traz desafios e impactos negativos para o mercado de trabalho. Um dos principais receios é a substituição de empregos tradicionais por automação, especialmente em setores que envolvem tarefas mecânicas e repetitivas. Isso pode levar a uma redução na demanda por mão de obra em determi-

nadas áreas, causando desemprego estrutural e aumentando a pressão por reskilling e upskilling da força de trabalho para se adaptar às novas demandas tecnológicas. Além disso, a crescente dependência da IA também levanta preocupações éticas, como a possibilidade de discriminação algorítmica e viés nos sistemas de recrutamento e seleção, o que pode perpetuar desigualdades sociais. Portanto, é crucial que a implementação da IA seja acompanhada de políticas públicas e estratégias empresariais que busquem mitigar esses impactos negativos, promovendo uma transição justa e equitativa para o novo cenário tecnológico.

Segundo pesquisa da McKinsey em 2018 – que teve como objetivo compreender a IA, o futuro do trabalho e o impacto da automação nas habilidades –, produtos e serviços com IA podem chegar a valer \$13 trilhões até 2030. Esse mercado futuro impulsionou países a liderar na IA, com investimentos vultosos, centros de pesquisa e integração da IA na economia e governo. Conscientes do impacto, regulamentam seu papel na sociedade. A IA pode aumentar distâncias econômicas entre líderes e demais países em até 25% e trazer lucros dobrados para empresas líderes em IA. Profissionais com IA terão mercado de trabalho maior e salários superiores aos de tarefas automatizadas (Bughin et al, 2018; Carvalho, 2021). Até 2030, a simulação média feita pelo instituto mostra que cerca de 70% das empresas podem ter adotado pelo menos um tipo de tecnologia de IA (Bughin et al, 2018).

Um desafio crucial é que a adoção da IA pode ampliar as disparidades entre países, empresas e/ou trabalhadores. A IA pode impulsionar a atividade econômica, mas a distribuição dos benefícios provavelmente será desigual. Em relação aos países, a IA pode aprofundar as lacunas entre as nações, reforçando a atual divisão digital. Os países podem precisar de estratégias e respostas diferentes devido às diferentes taxas de adoção da IA. As nações líderes no desenvolvimento e uso de IA podem capturar um adicional de 20 a 25 por cento em benefícios econômicos líquidos em comparação com hoje, enquanto os países em desenvolvimento podem capturar apenas cerca de 5 a 15 por cento (Bughin et al, 2018).

Sobre as empresas, as tecnologias de IA podem levar a uma lacuna de desempenho entre pioneiros de um lado e adotantes lentos e não adotantes do outro. No extremo do espectro, os pioneiros (empresas que absorvem totalmente ferramentas de IA em toda a sua organização ao longo dos próximos cinco a sete anos) provavelmente

se beneficiarão de forma desproporcional. Até 2030, eles poderiam potencialmente dobrar seu fluxo de caixa (benefício econômico capturado menos custos de investimento e transição associados), o que implica em crescimento adicional anual de fluxo de caixa líquido de cerca de 6 por cento por mais de uma década (Bughin et al, 2018).

Uma lacuna crescente também pode surgir no nível dos trabalhadores individuais. A demanda por empregos pode se afastar de tarefas repetitivas em direção a aquelas que são social e cognitivamente orientadas, além de outras que envolvem atividades difíceis de automatizar e requerem mais habilidades digitais. Perfis de emprego caracterizados por tarefas repetitivas e atividades que exigem baixas habilidades digitais podem experimentar o maior declínio como parte do emprego total, passando de cerca de 40 por cento para cerca de 30 por cento até 2030. O maior ganho na parcela pode estar em atividades não repetitivas e aquelas que exigem altas habilidades digitais, aumentando de cerca de 40 por cento para mais de 50 por cento. As consequências diretas dessa lacuna crescente no emprego e nos salários seriam uma intensificação da competição por pessoas, especialmente aquelas habilidosas no desenvolvimento e utilização de ferramentas de IA, e um excesso estrutural de oferta para uma parcela ainda relativamente alta de pessoas que não possuem as habilidades digitais e cognitivas necessárias para trabalhar com máquinas (Bughin et al, 2018).

A expansão da IA gera diversos efeitos secundários, como a criação de cursos de graduação em IA e a migração de pesquisadores de renome ou de estudantes de pós-graduação das universidades para grandes empresas ou para a criação de startups, muitas vezes antes de completarem seus cursos. Segundo Carvalho (2021), os estímulos à pesquisa e à tecnologia de IA no Brasil carecem de estratégia clara. O país está, segundo o autor, se contentando em ser um mero consumidor de produtos e serviços de IA. Há grandes preocupações sobre o impacto da IA no emprego com profissões que podem desaparecer, substituídas por atividades desconhecidas. Listas frequentemente destacam as profissões mais ameaçadas pela expansão da AI (Carvalho, 2021).

A simulação examinou sete possíveis vias de impacto da IA na economia. As três primeiras estão relacionadas ao impacto da adoção da IA na necessidade e na mistura dos fatores de produção que têm um impacto direto na produtividade das empresas. As outras quatro são externalidades relacionadas à adoção da IA e relacionadas ao amplo ambiente econômico e à transição para a IA. Em relação ao impacto econômico

direto da IA nas dimensões de produção, o instituto examinou, como sendo a primeira via, entradas complementares adicionais para melhorar a produtividade, o que os economistas chamam de “aumento” do trabalho e do capital. As entradas necessárias para operar uma nova capacidade de IA incluem novos engenheiros e analistas de big data que desenvolvem e implementam soluções de IA. Na segunda via, temos o investimento em tecnologias de IA economiza não apenas em mão de obra – dado que as máquinas passam a assumir as tarefas atualmente realizadas por humanos – mas também em capital antigo, permitindo, por exemplo, a manutenção preventiva que aumenta a vida útil dos equipamentos reduzindo, assim, a necessidade de investir em substituições. A terceira via implica na inovação de processos deve permitir que as empresas produzam o mesmo resultado com entradas menores, enquanto a inovação de produtos e serviços tende a impulsionar a produção e o nível de insumos (Bughin et al, 2018).

O uso de ferramentas e técnicas de IA pode contribuir para fluxos globais entre países e facilitar o comércio transfronteiriço mais eficiente. Nesse sentido, países mais conectados e mais participativos em fluxos globais claramente se beneficiariam mais da IA. Ganhos na atividade econômica podem ser reinvestidos e continuar a produzir crescimento. No entanto, a atividade econômica expandida também pode implicar externalidades negativas decorrentes de custos de transição da implementação de tecnologias de IA e custos estruturais adicionais relacionados à perda de competitividade nas empresas que não adotam a IA ou aos trabalhadores deslocados porque faltam as habilidades para operar em uma economia baseada em IA (Bughin et al, 2018).

Para traçar um quadro mais completo do impacto econômico da IA, o instituto apontou quatro dimensões adicionais, tanto positivas quanto negativas. Na quarta via, temos os ganhos econômicos a partir de fluxos de dados globais aumentados. Os dados digitais agora representam uma parcela maior do que no passado dos fluxos transfronteiriços internacionais na forma de troca de conhecimento e informações e transações diretas, como o comércio eletrônico transfronteiriço. O instituto simulou que a IA poderia representar até 20 por cento da contribuição dos fluxos de dados e digitais, ou um impacto de 1,5 por cento até 2030 em comparação com hoje (Bughin et al, 2018).

A quinta via diz da criação de riqueza e do reinvestimento. À medida que a IA contribui para a maior produtividade das economias, a produção aumentada a partir de ganhos de eficiência e inovações pode ser, em tese, repassada aos trabalhadores na forma de salários e aos empreendedores e empresas na forma de lucros. A geração de riqueza induzida pela IA poderia criar efeitos colaterais que impulsionam o crescimento econômico (Bughin et al, 2018).

Já a sexta via de impacto aborda os custos de transição e implementação. Uma série de custos provavelmente serão incorridos durante a execução da transição para a IA. As empresas provavelmente incorrerão em custos ao reestruturar suas organizações. Alguns trabalhadores podem ser substituídos por novas tecnologias, e as empresas podem precisar pagar custos associados, como indenização. As empresas também precisam atualizar as habilidades de seus trabalhadores existentes. À medida que muitas tarefas são automatizadas, os funcionários precisam se adaptar a novos tipos de trabalho, e muitos trabalhadores precisarão ser treinados para usar novas ferramentas digitais e de IA para suas operações diárias (Bughin et al, 2018).

A sétima via diz das externalidades negativas. À medida que as empresas adotam e absorvem a IA, a pressão sobre o emprego e os salários provavelmente aumentará, o que pode reduzir a parcela do trabalho na renda e no crescimento econômico potencial – ciclicamente por meio da perda de consumo durante períodos em que os indivíduos estão desempregados ou em requalificação, e estruturalmente por meio de um efeito de renda relativa. Os trabalhadores deslocados podem precisar fazer cursos de requalificação fornecidos e apoiados por governos e empresas se desejarem se reintegrar rapidamente na força de trabalho. Outro custo é o apoio do governo aos trabalhadores afetados na forma de benefícios de desemprego e outras provisões sociais (Bughin et al, 2018).

Em suma, a inserção da inteligência artificial (IA) no mercado de trabalho tem gerado um conjunto complexo de impactos, tanto positivos quanto negativos, que abrangem diversos aspectos da economia e sociedade. Os benefícios da IA incluem maior eficiência operacional, produtividade aprimorada e potencial para a criação de empregos especializados em tecnologias de IA. No entanto, o avanço tecnológico também suscita preocupações, como o deslocamento de trabalhadores em setores automatizados, acentuando as desigualdades econômicas e o risco de discriminação algorítmica.

A abordagem de políticas públicas e estratégias empresariais é crucial para mitigar esses impactos negativos e garantir uma transição justa para a era da IA. A implementação responsável da IA deve ser acompanhada por regulamentações que assegurem a privacidade dos dados, promovam a transparência e a equidade nas decisões algorítmicas e incentivem o reskilling e upskilling da força de trabalho. Além disso, é necessário um esforço conjunto para aprimorar a educação e a formação profissional, preparando os trabalhadores para as novas demandas tecnológicas e facilitando a inclusão digital.

A transformação impulsionada pela IA não se restringe apenas ao ambiente de trabalho, mas tem implicações mais amplas na economia global e nas relações internacionais. Países conectados e participativos em fluxos globais têm maior potencial para colher benefícios econômicos da IA, enquanto os desafios relacionados a desigualdades e custos de transição exigem uma abordagem colaborativa e coordenada entre governos, empresas e sociedade civil. Dessa forma, a incorporação da IA no mercado de trabalho não apenas exige adaptação, mas também oferece oportunidades para moldar um futuro mais equitativo e próspero, onde a tecnologia seja um instrumento de progresso para todos.

Considerações finais

A revolução industrial em curso, impulsionada pelo avanço da Inteligência Artificial (IA), redefiniu os paradigmas da sociedade contemporânea, estabelecendo um cenário complexo e interconectado que envolve avanços tecnológicos e desafios éticos. À medida que máquinas se elevam além das tarefas manuais, adentrando os domínios da inteligência e da tomada de decisões racionais, testemunhamos uma mudança de paradigma que transcende fronteiras disciplinares e geográficas. A convergência de fatores-chave, como a acessibilidade de processamento e memória, avanços em redes neurais profundas e a disponibilidade massiva de dados, intensificou o entusiasmo em torno das capacidades da IA. No entanto, esse entusiasmo é acompanhado por uma série de preocupações válidas, refletindo o desejo de se estabelecer um ambiente ético e responsável para a transformação em curso.

A IA, multifacetada em sua evolução, tem o poder de redefinir a interseção entre tecnologia e humanidade. Seja na resolução de tarefas complexas ou na emulação da inteligência humana, a IA revela sua capacidade de adaptação, moldando setores como diagnósticos médicos, educação e visão computacional. No entanto, esse progresso não ocorre sem dilemas éticos e sociais intrincados. A transparência nas decisões algorítmicas, a equidade na distribuição de benefícios e a salvaguarda da privacidade emergem como questões críticas que demandam soluções colaborativas e regulamentações criteriosas.

A complexidade da interação entre IA e ética vai além do desenvolvimento tecnológico, abraçando questões de privacidade, segurança e impactos socioeconômicos. A busca por algoritmos compreensíveis e justificáveis sustenta a confiança nas decisões algorítmicas, enquanto a legislação de proteção de dados ressalta a importância crescente da transparência e privacidade. A autonomia compartilhada e a compreensão dos impactos da IA na sociedade representam desafios globais que transcendem os limites das disciplinas tradicionais.

O diálogo em torno da ética na IA não é mera especulação acadêmica, mas uma necessidade premente. À medida que a influência da IA continua a se expandir, a responsabilidade de construir uma base ética sólida recai sobre a sociedade como um todo. O engajamento colaborativo entre diversos campos do conhecimento, governos, organizações e a sociedade civil é crucial para definir um caminho ético sustentável. Dessa colaboração emergirá um ambiente onde a IA, alinhada com valores humanos e justiça social, potencialize verdadeiramente o progresso e o bem-estar da humanidade.

No contexto do mercado de trabalho, a incorporação da IA traz consigo a dualidade entre potencialidades e desafios. Sua capacidade de aumentar a eficiência e a produtividade é inegável, liberando recursos para atividades mais criativas. Contudo, a automação também gera apreensões, com preocupações de desemprego e discriminação algorítmica. A implementação responsável da IA, acompanhada de políticas públicas e estratégias empresariais adequadas, é essencial para garantir uma transição justa e equitativa, além de impulsionar o desenvolvimento de habilidades relevantes para o novo paradigma.

Na arena global, a IA está tecendo uma rede complexa de impactos econômicos e sociais, estabelecendo um campo onde colaboração e coordenação internacional são

imperativos. Países que se engajam plenamente em fluxos globais têm o potencial de colher benefícios econômicos, mas o desafio das desigualdades e custos de transição exige uma abordagem conjunta e orientada para o bem comum.

Assim, a revolução da IA transcende os limites da tecnologia, redefinindo nossa sociedade em todas as suas facetas. É uma jornada de descoberta e responsabilidade, onde a interseção da inteligência artificial com a ética moldará o curso da humanidade. A busca por um equilíbrio entre inovação, ética e equidade é a essência do nosso futuro coletivo, onde a IA é mais do que um instrumento tecnológico, é um reflexo de nossos valores e aspirações.

Referências

- ABRAMOVAY, Ricardo. *O fim do trabalho: Entre a distopia e a emancipação*. Estudos Avançados, São Paulo, v. 35, n. 101, 2021.
- BERTALAN, V. G.; RUIZ, E. E. Predicting Judicial Outcomes in the Brazilian Legal System Using Textual Features. In: *DHandNLP@PROPOR*, 2020.
- BUGHIN, Jacques et al. *Notes from the AI frontier: Modeling the impact of AI on the world economy*. McKinsey Global Institute, 2018.
- CARVALHO, André Carlos Ponce de Leon Ferreira de. *Inteligência Artificial: Riscos, benefícios e uso responsável*. Estudos Avançados, São Paulo, v. 35, n. 101, 2021.
- COZMAN, Fabio Gagliardi. *No canal da Inteligência Artificial: Nova temporada de desgrehados e empertigados*. Estudos Avançados, São Paulo, v. 35, n. 101, 2021.
- DIETTERICH, T. G.; HORVITZ, E. Rise of concerns about AI: Reflections and directions. *Communications of the ACM*, v. 58, n. 10, p. 38-40, 2015.
- DEVLIN, Jacob et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, p. 4171-4186, Minneapolis, Minnesota, 2019.
- LUDERMIR, Teresa Bernarda. *Inteligência Artificial e Aprendizado de Máquina: Estado atual e tendências*. Estudos Avançados, São Paulo, v. 35, n. 101, 2021.
- NETO, Victo José da Silva; BONACELLI, Maria Beatriz Machado; PACHECO, Carlos Américo. O Sistema Tecnológico Digital: Inteligência Artificial, Computação em Nuvem e Big Data. *Revista Brasileira de Inovação*, Campinas, v. 19, e0200024, p. 1-31, 2020.
- SICHMAN, Jaime Simão. *Inteligência Artificial e sociedade: Avanços e riscos*. Estudos Avançados, São

Paulo, v. 35, n. 101, 2021.

VASWANI, A. *et al.* *Attention is all you need.* In: *Advances in Neural Information Processing Systems*, 2017, p. 5998--6008.

VICARI, Rosa Maria. *Influências das Tecnologias da Inteligência Artificial no ensino.* *Estudos Avançados, São Paulo, v. 35, n. 101, 2021.*





■ Giselle Beiguelman, frames do vídeo generativo da série Flora Rebellis, versão 2024.
Imagens geradas com Inteligência Artificial (Style Gans)

■ [Acesse o vídeo clicando na imagem](#)





INTELIGÊNCIA ARTIFICIAL, TECNOLOGIA E AS SERVENTIAS EXTRAJUDICIAIS

DOI: <https://doi.org/10.35699/2965-6931.2023.47716>

OSVALDO JOSÉ GONÇALVES DE MESQUITA FILHO*; MARCELO DE OLIVEIRA MILAGRES**

RESUMO: O artigo analisa a relação entre a inteligência artificial, a tecnologia e as serventias extrajudiciais, mais conhecidas como “cartórios”. No sistema adotado no Brasil, a atividade cartorária está presente em diversos atos da vida do brasileiro, desde o nascimento até o registro imobiliário. Tem-se o objetivo de revisitar o processo de modernização dos cartórios até o momento atual, em que se discute a utilização da inteligência artificial. É uma análise dogmática, em um primeiro momento, com a posterior avaliação crítica, examinando-se os desdobramentos e as possibilidades de aplicação da inteligência artificial no contexto da interface entre cidadãos e os serviços cartorários. Para cumprir o objetivo geral do artigo, as metodologias de análise de dados utilizadas se baseiam na vertente jurídico-social, com o método dedutivo como metodologia de pesquisa. Os resultados demonstram mais dúvidas do que respostas, principalmente pela novidade da temática. Conclui-se que o ideal será a conciliação entre o sistema notarial/registral e a inteligência artificial, de forma conjugada para que ambos evoluam.

PALAVRAS-CHAVE: inteligência artificial; serventias extrajudiciais; tecnologia; sistema notarial e registral; blockchain.

Artificial intelligence, technology and extrajudicial services

ABSTRACT The article analyzes the relationship between artificial intelligence, technology and notary/registral services, better known as “notaries” or “registries”. In the system adopted in Brazil, the notary/registral activity is present in several acts of Brazilian life, from birth to land registration. The aim is to resume the modernization process of this system up to the present moment, in which the use of artificial intelligence is discussed. It is a dogmatic analysis, at first, with subsequent criticism, analyzing the consequences and possibilities of applying artificial intelligence in this field. The results show more questions than answers, mainly due to the novelty of the theme. It is concluded that the ideal will be the conciliation between the notary/registral system and artificial intelligence, in a conjugated way so that both evolve. The methodological option is the legal-social aspect and the scientific reasoning is the deductive one.

KEYWORDS: artificial intelligence; notaries and registries; technology; notary and registral system; blockchain.

* Universidade Federal de Minas Gerais / ** Universidade Federal de Minas Gerais

Introdução

As serventias extrajudiciais ou, simplesmente, os “cartórios”, como são conhecidos pela maioria da população, sempre fizeram parte do cotidiano brasileiro. Atividades como reconhecer firma de uma assinatura; autenticar documentos; casar-se ou divorciar-se; registrar o nascimento de um(a) filho(a) ou um óbito de um familiar; celebrar escritura pública de compra e venda de um imóvel e, posteriormente, levar tal documento para o registro imobiliário; entre tantos outros exemplos desse cotidiano, perpassam a vida de cada cidadão, de seu nascimento ao óbito, mesmo que historicamente esses momentos sejam vistos como desnecessários, com longas filas e tempo de espera elevado. A fim de apresentar este percurso, a metodologia será analítica e descritiva, ou seja, irá perpassar, sucintamente, toda a trajetória, demonstrando de “onde viemos” e “para onde vamos”.

O objetivo geral deste artigo é mostrar a mudança desse cenário, demonstrando que as serventias extrajudiciais passaram – e estão passando – por um processo de modernização, inserindo, cada vez mais, a tecnologia no desempenho de suas atribuições. É uma construção que não é de hoje ainda se encontra em fase de implementação e de aperfeiçoamento.

A fim de apresentar tal percurso, a metodologia utilizada no artigo será analítica e descritiva, ou seja, irá perpassar, sucintamente, por toda a trajetória dos cartórios, demonstrando as origens e o que está por vir. Por certo, a discussão sobre o futuro ainda carrega mais perguntas do que respostas, especialmente no tocante à utilização da inteligência artificial no âmbito das serventias extrajudiciais, o que também será objeto de análise neste ensaio.

No primeiro tópico, será feita a apresentação das serventias extrajudiciais, de forma propedêutica e com o fim de demonstrar a construção – conceitual, histórica e sistêmica – da atividade notarial e registral. Além disso, discutir-se-á a importância dessa função para a sociedade, em seu viés social e econômico. No segundo tópico, iniciar-se-á a análise sobre a relação entre a tecnologia e a atividade delegada, em uma exposição dogmática sobre a temática. Por fim, adentrar-se-á na temática da inteligência artificial, buscando uma correlação com a atividade notarial e registral.

A opção metodológica do artigo, para fins de análise de dados, se configura na

vertente jurídico-social, uma vez que compreende o fenômeno jurídico no ambiente social mais amplo (Gustin; Dias; Nicácio, 2020, p. 65). Por sua vez, o raciocínio científico está amparado no método dedutivo, já que se parte de premissas gerais para, no decorrer do texto, tecer considerações específicas (Gustin; Dias; Nicácio, 2020, p. 71).

1 É o modelo característicos dos países socialistas, em que o notário se encontrava inserido na estrutura hierárquica e burocrática do Estado, sendo um mero funcionário ou empregado estatal. Caiu em desuso com o declínio dos países socialistas e encontra-se praticamente extinto. (Kumpel, 2017, p. 112)

A origem das serventias extrajudiciais no Brasil

Existem, no Brasil, mais de treze mil serventias extrajudiciais, presentes em todos os municípios brasileiros (CNJ, 2023). Os cartórios se configuram como uma das instituições mais longínquas e bem-estruturadas da ordem jurídico-social do país, presentes desde o Brasil-colônia (Tutikian, 2011), embora não se enquadrem como uma singularidade brasileira, já que se trata de instituto mundialmente arcaico, presente desde os primórdios romano-egípcios (Lago, 2022).

Na estrutura jurídica mundial, existem três classificações predominantes para se definir a organização notarial: o Notariado Administrativo¹, o Notariado Anglo-Saxão² e o Notariado Latino, tendo sido este último modelo o adotado pelo sistema brasileiro. O notário do tipo latino é um profissional do Direito, titular de função pública, nomeado pelo Estado para conferir autenticidade aos atos e negócios jurídicos contidos nos documentos que produz, e orientar e assessorar os usuários, com imparcialidade e independência (UINL, 2023).

O sistema do notariado latino, ao contrário do que o próprio nome pode sugerir, está presente na maior parte do mundo e não somente nos países de língua latina. É adotado em mais de 120 países, abrangendo dois terços da população mundial e mais de 60% do Produto Interno Bruto (PIB) do planeta (UINL, 2023). A seguir, será feita uma análise – propedêutica e sucinta, como não poderia ser diferente – da estrutura desse sistema, suas principais características e princípios, com alguns pontos controvertidos na doutrina e jurisprudência.

Na sistemática latina, a atividade notarial/registral não é de base judicial, nem administrativa, pois se trata de profissional do Direito que exerce em caráter privado e com independência a função pública delegada pelo Estado, e, por isso, por isso, não

2 No sistema anglo-saxão, a figura do notary se limita a identificar os subscritores do documento, a reconhecer a assinatura, a apor o respectivo selo e assinatura no documento, sem receber ou interpretar qualquer declaração de vontade dos particulares. É o modelo vigente na common law, em que a oralidade assume um papel muito mais importante, já que as leis escritas possuem função secundária, apenas auxiliando a esclarecer o costume. Desta feita, o papel do notário e registrador é bem menos importante, podendo tal função ser exercida por particulares, sem necessidade de formação jurídica, sendo que a intervenção notarial é, na maioria das vezes, superveniente à existência do documento. (Kumpel, 2017, p. 110-111)

3 Art. 236 da Constituição da República de 1988.

4 Cf. ADI MC nffl 1.926/PE (Dje. 10 set. 1999); ADI MC nffl 1.378/ES (Dje. 30 maio. 1997); ADI MC nffl 1.444/PR Dje 29 ago. 1997); ADI nffl 1.709/MT.

5 Há divergência doutrinária nesse sentido, mas se optou por trazer a posição majoritária e vigente no atual ordenamento jurídico, conforme artigo 22 da Lei nffl 8.935/1994.

integra a estrutura estatal (Rodrigues, 2022, p. 858). Ora regida por normas de Direito Público, ora por normas de Direito Privado. Dip (apud Loureiro, 2019, p. 56) conceitua essa dualidade como “binômio-tensivo” da atividade notarial e registral, o que demanda uma qualificação jurídica ainda maior do titular da serventia.

No ordenamento jurídico brasileiro, à semelhança da organização latina, as atividades notariais e registrais são funções públicas, próprias do Estado, mas exercidas em caráter privado e mediante delegação estatal³. O delegatário possui total autonomia organizacional, inexistindo relação de hierarquia com o Estado, e, sim, de fiscalização, exercida por meio do Poder Judiciário.

A natureza jurídica dos notários e registradores era controvertida na doutrina e na jurisprudência brasileiras. Nos últimos anos, sedimentou-se que os notários e registradores não são servidores públicos, mas, sim, particulares em colaboração com o Estado, prestadores de um serviço público que receberam mediante delegação.

Pelo fato de não serem servidores públicos no sentido estrito, são independentes na prestação dos seus serviços, responsáveis pelo gerenciamento técnico-administrativo da serventia extrajudicial e remunerados exclusivamente pelos emolumentos, valores pagos pelos usuários e que tem natureza jurídica de taxa⁴. Em caso de falha na prestação do serviço, o cidadão pode postular indenização em face do titular do cartório, que tem responsabilidade subjetiva⁵, e/ou do Estado, que tem responsabilidade objetiva e direta, o que justifica pela natureza pública da função exercida.

Quando se fala que a atividade notarial e registral apresenta características empresariais, refere-se ao fato de que o delegatário é o responsável por toda a gestão da serventia, com seu gerenciamento técnico, administrativo e financeiro. Por sua vez, Ceneviva defende que a função não deve visar ao lucro, uma vez que deveria servir ao interesse público, e, portanto, não poderia ser comparada à atividade empresarial (Ceneviva, 2002, p. 145).

É certo que a finalidade é pública, mas para a consecução de tal fim, os meios – organização técnica, administrativa e financeira da serventia – são semelhantes ao de uma empresa, o que justifica tal comparação. Dizer que se assemelha ao modelo empresarial não é negar a finalidade pública, mas, sim, buscar que seja cumprida de forma ainda mais eficiente. Compreender a função notarial/registral dessa forma é o desejável para o seu avanço e a modernização do serviço.

O aspecto publicista da atividade notarial e registral está ligado à sua própria natureza, que é de serviço público. Apesar de ter havido a delegação estatal, perdura a finalidade do interesse público, haja vista a função ser pública, de grande importância para a sociedade e fazer parte da estrutura estatal.

Esse contexto justifica que o Estado continue exercendo a irrenunciável função de controle, a qual se dá em dois níveis: passivo, que se dá pela informação e fiscalização, a ser realizada pelo Poder Judiciário; e ativo, por meio da regulamentação normativa, a qual deve ser obrigatoriamente seguida pelo delegatário (Kindel, 2007, p. 66). Quanto à regulação e controle estatal perante a atividade notarial e registral, vale destacar o seguinte trecho de Ribeiro (2009, p. 80):

[...] a atividade notarial e de registro é uma atividade jurídica que está, no Estado, muito próxima da atividade jurisdicional. Embora não se desenvolva para a solução de litígios, como a atividade jurisdicional pura, ela é uma atividade pré-contenciosa, e *deve ser utilizada com um regramento público eficaz, com uma regulação forte do Estado e uma auto-regulação privada. Deve ser utilizada em benefício dos cidadãos.* (grifos nossos).

Ribeiro defende que a fiscalização não basta, devendo haver a regulação, que é mais ampla e envolve a elaboração de normas pertinentes, a implementação/efetivação desse ordenamento e, por fim, a fiscalização/punição de eventuais infrações (Ribeiro, 2009, p. 86-95). O Estado, ao se desonerar da execução direta e exclusiva desse serviço público, assume um dever de intervenção para que a atividade seja bem prestada, garantindo a sua adaptabilidade e atualização em face da evolução sociojurídica.

No cenário brasileiro, essa regulação, tão defendida por Ribeiro, dá-se, sobretudo, pela atuação do Conselho Nacional de Justiça (CNJ), que é responsável pelas principais normativas relacionadas à atividade notarial e registral. Cabe destacar, ainda, as Corregedorias de Justiça Estaduais, que, além de fiscalizar as serventias extrajudiciais, editam as normas locais.

A atuação do CNJ e das Corregedorias de Justiça Estaduais foi essencial para a modernização da atividade cartorária, que assumiram um papel de vanguarda e, a partir das suas normativas, implementaram uma série de inovações mesmo antes da regulação da temática a nível de lei federal.

A função de controle estatal também se verifica na organização estrutural do sistema notarial e registral, característica típica dos chamados notariados numerários, nos

6 Para aprofundamento na temática dos princípios e o direito notarial e registral, sugere-se: Kern; Costa Junior, 2022.

7 A excepcionalidade fica por conta dos loteamentos imobiliários e das incorporações imobiliárias, em que é permitido o acesso aos documentos referentes aos empreendimentos de forma direta, independentemente de certidão.

8 Excepcionalmente, pode ser ativa, tal como nas notificações realizadas pelos notários/registradores para dar ciência ao notificado sobre algum registro, averbação ou procedimento que lhe diga respeito.

quais o Estado deve determinar o número de serventias existentes e suas principais características. O Supremo Tribunal Federal (STF), por meio da Ação Direta de Inconstitucionalidade (ADI) nº 4.140/GO, definiu que é responsabilidade do ente federativo estadual: a definição e organização das serventias; a criação ou extinção dos cartórios; e a definição do número de serventias em determinado município e, como um todo, no estado.

Com relação aos princípios norteadores da atividade, serão apresentados apenas dois – a publicidade e a rogação – por entender-se que são os de maior correlação com o processo de modernização da função. Ressaltando que não há qualquer pretensão de prolongar-se na temática⁶.

O primeiro deles, como não poderia ser diferente, é o princípio da publicidade, fundamento basilar da atividade notarial e registral. Para Lopes, a publicidade transcende a principiologia, tratando-se de utilidade jurídico-social, publicizando e/ou criando relações jurídicas e direitos (Lopes, 1997, p. 18).

A publicidade é universal, destinando-se a toda e qualquer pessoa, independentemente de interesse direito ou indireto (Ceneviva, 1979, p. 68). Em regra, é indireta, sendo obtida por meio de certidão⁷; e passiva, já que o interessado que deve procurar o serviço para obter alguma informação⁸. Com a incorporação da tecnologia à atividade delegada, a publicidade se tornou mais ampla e acessível, alargando, ainda mais, a importância desse princípio dos registros (Almeida, 1966).

Vale destacar, ainda, o princípio da rogação ou da instância, que determina que os delegatários não podem atuar de ofício. Sem solicitação da parte, o notário e o registrador não podem praticar seus atos, devendo haver um requerimento – oral ou escrito – para a sua realização. A forma como se dá essa postulação se alterou consideravelmente com a modernização das serventias extrajudiciais, que hoje pode ser totalmente digital, extirpando a necessidade das “idas ao cartório”, metáfora mencionada na introdução.

Passada a apresentação inicial da atividade cartorária, é imprescindível a análise da função social e econômica dos notários e registradores, demonstrando a sua importância na sociedade. É imperioso realizar tal exposição, já que a função notarial/registral é, por muitas vezes, rotulada – com demasiada pressa e irrefletido descaso – como ultrapassada velharia formal (Campilongo, 2014, p. 108) e como instituição existente somente no Brasil, o que já se demonstrou falso.

Segundo Campilongo (2014, p. 17), a função social do notariado latino é sedimentada em três pressupostos: a criação de ambiente propício à produção de eficiência econômica; a estabilização da confiança como mecanismo de redução da complexidade; e o acesso à informação como garantia de legalidade, transparência e, principalmente, imparcialidade. Tais premissas – eficiência, confiança e imparcialidade – seriam reunidas a fim de garantir um cenário de menos litigiosidade, de mais credibilidade nas instituições, com as relações jurídicas pautadas pelo equilíbrio e mediação de terceiro imparcial, o que é propício ao desenvolvimento econômico e social.

A confiança é, portanto, mecanismo basilar da atividade cartorária, sendo, ao mesmo tempo, paradoxal, visto que as partes confiam porque desconfiam (Campilongo, 2014, p. 22). Tal cenário é normal nos dias de hoje, em que a dinâmica empresarial comumente envolve partes que não se conhecem, originárias de locais diferentes no mundo. A atuação imparcial do notário, colocando esses sujeitos em relação de equidistância, provê confiança e segurança negocial.

Para arrematar, é necessário colacionar trecho de Campilongo (2014, p. 93):

A institucionalização da confiança reduz custos de transação, mitiga assimetrias de informação, internaliza externalidades positivas na produção de escrituras e contratos adequados. No ambiente institucional brasileiro, trata-se de regulação que aproxima os mercados de resultados pretendidos pelo paradigma competitivo. O Estado e o Direito, na análise da regulação referente aos notários, não podem ignorar os efeitos percebidos pelos mercados. (grifos nossos).

No ideário comum, questionam-se os custos da atividade notarial e registral, argumentando-se que seriam elevados face à importância diminuta do serviço. Arruñada (1995) demonstra que o custo do notariado latino é muito baixo, especialmente se comparado aos países do sistema anglo-saxão. A atividade notarial/registral, nesses países, é de importância/atuação mínima, o que gera uma falta de segurança jurídica e o aumento dos custos de transação, especialmente pela atuação obrigatória das seguradoras (Faria; Lacerda, 2022).

Há, ainda, o aspecto de prevenção de litígios e de desjudicialização inerente à atividade notarial e registral. Cernelutti (apud Campilongo, 2014, p. 22) define bem o

9 A crise do subprime é o nome que foi dado ao ponto de mudança na economia e na cultura. Isto é, em seu centro, o resultado da bolha especulativa no mercado imobiliário que começou a estourar nos Estados Unidos em 2006 e tem causado rupturas ao longo de vários países em termos de crises financeiras e quebra do crédito global. (Shiller, 2008, p. 01)

papel que se espera do notário: “quanto mais notário, menos juiz”. O caráter ex ante de intermediação imparcial, antes de qualquer litígio, é o desejável na sociedade atual. Sobre a temática, irretocável a conclusão de Rodrigues (2022, p. 863):

[...] o direito não pode, nem deve, ser analisado apenas sob a ótica da resolução de litígios. *Mais importante que resolvê-los é, na medida do possível, evitá-los, pacificando conflitos e assim eliminando germes de futuras demandas.* O litígio é oneroso, sua resolução é demorada e nem sempre tem o condão de encerrar o conflito, pois as partes, mesmo após a intervenção do Estado-juiz, com frequência não se pacificam, com a solução alcançada. Por isso que deve ser visto o litígio e examinada a possibilidade de resolução judicial como exceção, jamais como regra.

Por fim, destaca-se a teoria de Shiller, ganhador do Prêmio Nobel de Economia, em que o economista, ao analisar a crise do *subprime*⁹, defende que a implantação do notariado latino seria uma das soluções de longo prazo a serem adotadas nos Estados Unidos. Fundamenta a sua defesa no papel da atividade notarial para a democratização da informação e a correta instrumentalização/utilização das hipotecas (Shiller, 2008, p. 134).

Diante do avanço tecnológico inerente à sociedade atual, surgiu a necessidade – ou, talvez, a obrigatoriedade – das serventias extrajudiciais se adaptarem a esta nova realidade. Tal processo de modernização será explorado a seguir.

A tecnologia e as serventias extrajudiciais

Viver na era digital é uma epopeia desafiadora (Nalini, 2021, p. 456). Ser adaptativo e não se tornar obsoleto torna esse desafio ainda maior, raciocínio aplicável a todos os campos da vida pessoal e profissional. Não poderia ser diferente com a atividade notarial e registral, que teve de se adaptar – e ainda está se adaptando – à tecnologia e ao mundo moderno.

Schwab (2016) afirma que vivemos a Quarta Revolução Industrial, imersos em uma revolução tecnológica que transformou e continua a transformar a forma como vivemos, trabalhamos e nos relacionamentos. Diante desse cenário, surgiu a inquietação: como conciliar o desenvolvimento tecnológico e a função desempenhada pelas serventias extrajudiciais? Como garantir a segurança jurídica inerente à função e, ao mesmo tempo, modernizar-se para acompanhar os novos tempos?

Emergiu-se, então, a necessidade de modernização do arcabouço legislativo, de forma a regular a adaptação da atividade à tecnologia. Em um primeiro momento, a contudo, a legislação federal não atuou nesse sentido, cabendo às Corregedorias Estaduais e, principalmente, ao CNJ o papel de agente regulador dessas mudanças. O papel da normativa infralegal – provimentos, resoluções, recomendações e consolidações normativas estaduais – foi essencial para a regulação e o desenvolvimento da temática.

É importante ressaltar que inovar e modernizar não significa repudiar a construção histórica de um sistema tão consolidado e com a função socioeconômica que lhe é inerente. Nalini (2021, p. 468) afirma que o passado não precisa ser destruído, mas reintegrado às novas tecnologias, devendo-se remodelar as mentalidades e as práticas. Nessa mesma linha, Gállico (2021, p. 266) afirma que as serventias extrajudiciais devem se adaptar à velocidade da sociedade moderna, incorporando novas tecnologias, mas sempre sem perda da segurança jurídica.

Atuando nesse sentido, surgiram as primeiras normativas relativas à modernização da atividade notarial e registral. Restringe-se o escopo do artigo às normativas editadas pelo CNJ e à legislação federal, deixando-se de mencionar as regras editadas pelas Corregedorias Estaduais, pela própria dificuldade e complexidade da análise a ser realizada no âmbito nacional.

O ponto de partida foi a Recomendação nº 14/2014 do CNJ, que introduziu a ideia do Sistema de Registro de Imóveis Eletrônico (SREI). Recomendou-se a adoção de práticas voltadas ao registro eletrônico e às centrais eletrônicas, de forma a chamar atenção para a temática. Em seguida, o CNJ editou o Provimento nº 47/2015, que estabeleceu as primeiras diretrizes desse sistema eletrônico de registro imobiliário.

A partir de tais normativas, emergiram as regulamentações a nível estadual pelas respectivas Corregedorias, o que foi dando amplitude à temática. Concomitantemente, as entidades de classe se organizaram na criação, regulamentação e operacionalização

¹⁰ Destaque para a vanguardista Central de Informações do Registro Civil (CRC), regulamentada nacionalmente pelo Provimento nº 46/2015, em que é possível a solicitação e a execução de atos no Registro Civil das Pessoas Naturais.

¹¹ Art. 76 da Lei nº 13.465/2017.

¹² Art. 76, §4º, da Lei nº 13.465/2017.

¹³ Art. 76, §9º, da Lei nº 13.465/2017, que foi introduzido pela Lei nº 14.118/2021.

das centrais eletrônicas, nas quais passou a ser possível a solicitação e, até mesmo, a realização on-line de alguns serviços notariais e registrares¹⁰.

No momento em que essas medidas de modernização fervilhavam, emergiu a Lei nº 13.465/2017 – de âmbito federal e com natureza de lei ordinária –, que trouxe mais alguns avanços para a temática. A legislação trouxe expressamente o SREI, o qual seria implementado e operado, em âmbito nacional, pelo Operador Nacional do Sistema de Registro Eletrônico de Imóveis (ONR)¹¹.

A regulamentação da temática foi postergada, deixada a cargo do CNJ, que foi eleito como agente regulador do ONR e responsável pelo cumprimento do seu estatuto¹². É importante destacar essa legislação, que, a despeito das suas lacunas e da necessidade de regulamentação futura, foi paradigmática ao tratar expressamente da temática a nível federal.

De modo a regulamentar a SREI e o ONR, foi editado o Provimento nº 89/2019 pelo CNJ, que também revogou o anterior Provimento nº 47/2015. Cabe destacar alguns pontos da norma: a regulamentação do Código Nacional de Matrículas, que é uma sequência numérica de individualização do imóvel a nível nacional; as diretrizes para o SREI, regulamentando seus aspectos gerais; e a regulação do Serviço de Atendimento Eletrônico Compartilhado (SAEC), uma espécie de “balcão virtual”, destinado ao atendimento remoto dos usuários interligado a todas as serventias de registro de imóveis do país.

Cabe mencionar, ainda, que o Provimento nº 89/2019 estabeleceu as primeiras diretrizes acerca da ONR. Posteriormente, em abril de 2020, ocorreu Assembleia Eletrônica na qual foram aprovados os estatutos do órgão e foi eleita a primeira Diretoria Executiva. Por fim, o CNJ editou o Provimento nº 109/2020, que disciplinou a atuação do próprio órgão enquanto agente regulador da ONR.

É importante destacar a celeuma que existia em relação ao custeio da implementação e manutenção da SREI, que foi encerrada com a criação de um fundo, gerido pela ONR e subvencionado pelas serventias de Registro de Imóveis¹³. Adotou-se, então, a política de subsídios cruzados, com o recolhimento de valores por todos os registradores mobiliários na proporção da arrecadação e do número de atos praticados na serventia, o que ficou normatizado pelo Provimento nº 115/2021 do CNJ.

Vale mencionar, ainda, o recente Provimento nº 143/2023, responsável por re-

gular a estrutura, geração e validação do Código Nacional de Matrícula. Destaca-se o prazo de 1 (um) ano para a implantação em todas as matrículas imobiliárias do país¹⁴, que, caso seja cumprido, representará um grande avanço para a atividade registral.

¹⁴ Art. 13, III, do Provimento nffl 143/2023.

Toda essa adaptação, modernização e incorporação da tecnologia aos serviços notariais e registrais teve um grande catalisador: a pandemia do COVID-19, iniciada em meados do ano de 2020. Vivenciávamos o medo difuso do invisível, o temor havia virado rotina, estávamos longe de ter o controle de tudo.

¹⁵ Para tal temática, sugere-se: Chezzi, 2021.

Contudo, a vida em sociedade é dinâmica. A despeito do isolamento social, as relações econômicas/comerciais persistiram. A construção civil e os negócios imobiliários continuaram; o comércio teve de se reinventar, alavancando as vendas virtuais (e-commerce); o agronegócio persistiu bravamente, sendo base da economia brasileira à época. Enfim, a vida, mesmo que modificada e atribulada, teve de continuar e continuou. Como não poderia ser diferente, as serventias extrajudiciais também continuaram, sem interromper a prestação do serviço público.

Nesse período, tivemos a edição de diversas normativas pelo CNJ, algumas já mencionadas, mas o grande destaque ficou por conta do Provimento nffl 100/2020, em que se regulamentou a prática dos notariais de forma eletrônica. Tal mudança foi emblemática e disruptiva, o que se justifica até pela natureza da função do tabelião de notas, marcado por uma maior proximidade e contato entre usuário e delegatário. Em um dia, cidadão e tabelião estavam lado a lado no momento da prática do ato; outrora, passa-se a permitir a lavratura de forma totalmente eletrônica e virtual.

Permitiu-se, então, a formalização dos atos de forma totalmente eletrônica, com as partes assinando por meio de certificado digital e com a realização de uma videoconferência. Por meio de uma plataforma de comunicação *online*, é feita uma espécie de audiência, a ser obrigatoriamente gravada, realizando-se os trâmites a semelhança do que ocorre presencialmente na serventia extrajudicial.

Não se analisará a fundo as diversas mudanças ocorridas com a edição do Provimento nffl 100/2020¹⁵, cabendo destacar o seu papel de destaque na modernização das serventias extrajudiciais como um todo. O Brasil assumiu posição de vanguarda, até mesmo em relação aos países da União Europeia, os quais, na sua maioria, não têm o sistema notarial e de registro imobiliário de forma totalmente eletrônica (Patrão, 2022, p. 27).

¹⁶ Art. 23, §4º, da Lei nº 13.709/2018.

Recentemente, foi editada a Lei nº 14.382/2022, que regulamenta uma série de inovações em âmbito notarial e registral. Tais foram as mudanças que lhe renderam a alcunha de “Cartório Digital” por Rodrigues (2022, p. 137-156). As alterações tiveram denominadores comuns: a simplificação na prática dos atos, com vistas à modernização e à celeridade; e a regulação federal de normativas esparsas, que já vinham sendo previstas em normativas do CNJ e das Corregedorias Estaduais.

Há que se falar, ainda, no aparente conflito entre a Lei Geral de Proteção de Dados (LGPD) e a publicidade inerente à função cartorária. As serventias extrajudiciais tratam inúmeros dados, diariamente, encontrando-se no rol dos agentes públicos que devem adotar cuidados/mecanismos para o correto tratamento desses dados¹⁶. Por outro lado, a publicidade da função é universal e, em regra, não há que se obstaculizar o acesso aos dados constantes em seus arquivos.

O paradigma a ser adotado é o de Chezzi, em que se considera tal cenário como não conflituoso, bastando que haja parcimônia, cautela e rigor no tratamento dos dados e, doutro lado, não se obstaculize o acesso ao acervo registral/notarial a não ser em casos excepcionais. Vale destacar o seguinte trecho do autor:

A LGPD serve, portanto, à proteção do acervo integral e ao redimensionamento do papel das informações pessoais neste banco público de dados imobiliários, no caso do Registro de Imóveis. Presta-se a incrementar a segurança jurídica no prestígio da finalidade de suas funções e proporcionar o fluxo adequado de dados pessoais na era da informação, melhorando o desempenho dos serviços de registros públicos em prol de toda a sociedade (Chezzi, 2021, p. 149).

Com vistas a regulamentar a temática e pacificar eventuais celeumas, o CNJ editou o Provimento nº 134, em 2022. Resumidamente, regulamentou-se a forma como os notários e registradores irão tratar os dados, estabelecendo as medidas a serem adotadas para a correta adequação à LGPD.

Toda essa exposição analítica e descritiva foi essencial para se entender “aonde chegamos”, mas, diante do dinamismo social e da evolução tecnológica, as mudanças não param e, conseqüentemente, a modernização também deve continuar. As questões práticas e os debates doutrinários sobre a temática estão fervilhando, trazendo-se, a seguir, alguns assuntos em torno da inteligência artificial e as serventias extrajudiciais.

2.2 A inteligência artificial e as serventias extrajudiciais: “Para onde vamos”

Trabalhar a correlação entre a inteligência artificial e as serventias extrajudiciais não é um mero exercício de futurologia, visto que se trata de tema atual e baseado em evidências – práticas e doutrinárias. O tópico se centra, então, em três temas principais: a inteligência artificial, o fenômeno da tokenização e o *blockchain*,

O ponto de partida para a análise é o Provimento nº 038/2021 da Corregedoria-Geral de Justiça do Tribunal do Rio Grande do Sul, em que se regulamentou a lavratura de escrituras públicas de bens imóveis com o pagamento/contrapartida em tokens/criptoativos, possibilitando também o subsequente registro. De forma vanguardista, o tribunal gaúcho trouxe a temática da tokenização de imóveis e a utilização das moedas virtuais para dentro das serventias extrajudiciais, o que, por certo, é suscetível de elogios, mas também de dúvidas e críticas.

É dada a chancela do Direito para a utilização dessas novas tecnologias em âmbito imobiliário, mas de forma regulada e diante do olhar do Estado, neste caso representado pelos delegatários. Richter e Philippi (2022, p. 192) defendem que a utilização dessas modernidades é inevitável, mas apontam a contradição existente entre a atual regulação gaúcha e a própria razão de existir dessas tecnologias:

[...] como um direito, fundado na lei e no Direito, pode ser assegurado e legitimado por uma tecnologia orientada pelo não direito? Como seria possível legitimar direitos através de mineradores aleatórios via solução matemática? Qual é a garantia que o titular de ativos imobiliários digitais “token imobiliários” incidentes sobre imóvel físico tem, se o direito de propriedade está inscrito no Registro Imobiliário e os *tokens* de *blockchain* não tem qualquer conexão com eles?

A *ratio* da *blockchain* e de tecnologias emergentes é a fuga do aparato estatal, utilizando-se de uma rede de transações desregulada e sem interferência do Estado. Buscar a regulamentação da temática é imprescindível, mas deve-se compreender – com o auxílio de especialistas da área e o inevitável estudo verticalizado da temática – como tais tecnologias funcionam.

As críticas à regulamentação gaúcha são diversas, especialmente no tocante à relação estatal e do Direito com tais modernidades. Jacomino e Unger (2023) sustentam que a tokenização não representa mais do que um simples direito de crédito, não havendo que se falar em aquisição da propriedade por tal meio, com críticas à normativa gaúcha. Vale destacar os seguintes trechos dos autores (Jacomino; Unger, 2023, p. 8-10):

A constituição, alteração, extinção dos direitos reais se dá pela inscrição no Registro de Imóveis competente (art. 1.227 do CC). Mesmo nos casos em que a aquisição do direito real se dê *extra tabula*, nas estritas hipóteses legais, a plena disponibilidade do direito somente se alcança com o registro do título no ofício imobiliário competente.

[...] Não pode uma entidade privada, à míngua de qualquer disposição legal, arvorar-se em entidade registradora, promovendo e sacramentando intercâmbios econômicos com o fim de constituir direitos reais sobre uma “propriedade digital” decalcada de um ativo físico tangível – como é o bem imóvel. Não é possível agravar e radicalizar um raciocínio analógico para emular, no âmbito estrito de relações tipicamente obrigacionais, os atributos próprios de direitos reais. Não se criam direitos reais pela simples manifestação de vontade das partes.

Nesse mesmo sentido, também criticando a relação entre o Estado e a tecnologia do *blockchain*, assim dispõe Rodrigues (2021, p. 219-240):

Portanto, a discussão deve ser posicionada em seus reais termos: em que medida está a sociedade disposta a descartar agentes estatais imparciais, independentes e confiáveis para adotar terceiros privados? E a questão reflexa: em que medida está o Estado disposto a ceder parte do espaço que ocupa hoje e que pode ocupar amanhã em favor de agentes puramente privados?

Maziteli Neto e Brandelli (2019) ponderam que, embora a tecnologia *blockchain* permita a tokenização do direito de propriedade, isso não altera o direito real que a sustenta, posto que a propriedade imobiliária tem tratamento regulado na legislação civil e registral. Por fim, destaca-se a crítica de Jacomino e Unger (2023, p. 11), para os quais faltarão instrumentos legais-processuais para que os titulares da dita “propriedade digital” possam opor e defender seus direitos em face de quaisquer terceiros, inclusive o próprio Estado.

É importante ter parcimônia e buscar o estudo verticalizado da temática antes de qualquer regulamentação do assunto, sob pena de se abrir margem à insegurança jurídica. As iniciativas de “tokenizar” imóveis, sem a intermediação dos *gatekeepers* reconhecidos em lei – *in casu*, o serviço notarial e registral –, abrem espaço para fraudes envolvendo a possibilidade de duplo registro, com múltiplas inscrições/tokenizações do mesmo imóvel em diferentes *blockchains* e a possibilidade de negociação de ativos sem lastro (Jacomino; Unger, 2023, p. 08).

É inegável que tais tecnologias devem ser implementadas à prática imobiliária, mas é imprescindível que o aparato jurídico esteja ao seu lado, de forma a garantir a segurança jurídica como um todo. Coaduna-se com o posicionamento de Werbach (2018), ao defender que o sistema da *blockchain* depende do aparato jurídico-legal e deve, portanto, ser pensada em conjunto com o arcabouço normativo (Werbach, 2018, p. 552)¹⁷.

A inteligência artificial está por toda parte, não sendo diferente no ramo jurídico. No caso brasileiro, é uma realidade crescente no Processo Civil¹⁸, na execução fiscal¹⁹, no chamado Direito Digital²⁰ e, como não poderia deixar de ser, na seara notarial e registral. Neste tocante, a grande questão que surge é a seguinte: como conciliar a atividade das serventias extrajudiciais e a inteligência artificial? De um lado, está o serviço milenar e pautado na segurança jurídica; doutro, encontra-se a inteligência artificial, a automatização de práticas e os inúmeros avanços correlatos.

Neste ensaio, o que se defende não são obstáculos à implementação dessas tecnologias, mas que haja uma conciliação e adaptação ao aparato jurídico-normativo existente. Vejamos algumas ideias, apontadas como centrais para essa linha de argumentação.

Gálligo (2021, p. 267) defende a incorporação do *blockchain* e da inteligência artificial como técnica de armazenamento de dados e garantia da inalterabilidade do trato sucessivo, da mesma forma que os registros sempre incorporaram novas tecnologias. Miranda (2017, p. 92), por sua vez, aponta as vantagens da utilização da computação cognitiva na qualificação registral, caminhando na ideia de uma qualificação híbrida:

A qualificação realizada pelo sistema de computação cognitiva apresenta diversas vantagens: maior velocidade na leitura, interpretação e qualificação; custo reduzido; leitura integral e conferência de ponto por ponto, evitando-se leitura superficial; memória completa quanto aos requisitos e verificações a serem feitas; inexistências de erros por cansaço; facilidade na realização de verificações rotineiras; emissão de relatórios completos, entre outros.

A implementação da computação cognitiva²¹ representaria uma elevação de qualidade e eficiência na prestação do serviço, tratando-se de redução de custo e aperfeiçoamento do desempenho (Miranda, 2017, p. 83-84). Miranda (2017) defende o desempenho da tarefa de maneira híbrida – pela inteligência artificial, sistemas informatizados e pelos seres humanos –, com a conferência e aval sempre realizados pelo homem.

¹⁷ These are still early days for the blockchain (...). As big as the market has grown, there is far less at stake, and therefore far less path dependence, than there will be in three, or five, or ten years. Now is the time to develop hybrids of law and code. Regulators, legislators, and courts can take the initiative to create both clarity and explicit spaces for experimentation. (...) Like the Internet, the blockchain is a foundational technology, whose impacts could reach into every corner of the world. To move forward, though, law and distributed ledgers need each other. (Werbach, 2018, p. 552)

¹⁸ Por todos, Cf. Nunes; Lucon; Wolkart, 2020; e Vale; Pereira, 2023.

¹⁹ Por todos, Cf. Vale; Mazzola, 2022.

²⁰ Por todos, Cf. Barbosa et al., 2021.

²¹ Computação cognitiva é o nome que se dá ao conjunto de procedimentos, rotinas e sub-rotinas, software e hardware que consegue aproximar o desempenho de um sistema computacional à forma de processamento de informações de um ser humano. A principal vantagem de um sistema de computação cognitiva é a possibilidade de compreensão, pelo computador, de uma nova forma de conteúdo: conteúdo criado, organizado e estruturado para compreensão por seres humanos. (Miranda, 2017, p. 82)

22 Para que se compreenda a dita qualificação híbrida, assim como o procedimento sugerido pelo autor, sugere-se: Miranda, 2017, p. 84-99.

Alinha-se ao posicionamento acima, corroborado pelas ideias de Brandelli, defensor do uso conjunto de inteligências – a artificial e a humana. Para o autor, o nosso sistema de registro de direitos demanda uma qualificação jurídica importante, que é o cerne da atividade registral e que justifica a sua classificação como atividade intelectual (Brandelli, 2021, p. 396). Enquanto função intelectual, não há que se falar em substituição pela tecnologia, mas tão somente em conjugação, complementação e incorporação.

Com relação à tokenização de imóveis e a utilização do *blockchain*, há poucas respostas sedimentadas na prática notarial e registral, assim como na doutrina e na jurisprudência. Por sua vez, Jacomino e Unger (2023, p. 09-10) sintetizam as principais dúvidas existentes sobre a temática:

Deve-se abrir esta discussão aclarando-se as seguintes questões: a consumação do registro constitutivo da propriedade digital, nesta peculiar coordenação entre a matrícula e a *blockchain*, é uma atividade pública ou privada? O ato de registro da propriedade digital estará revestido de oficialidade e estatalidade – como reclamam a legislação civil e registral? Os direitos constituídos na plataforma eletrônica poderão ser exercidos *erga omnes*? Esses mesmos direitos serão tutelados e garantidos pelos tribunais, como ocorre com os típicos direitos e ações reais?

González e Fernández (2020, p. 344) argumentam – com base na experiência do Direito Comparado, especialmente o espanhol – pela necessidade de procedimentos e instituições que identifiquem fisicamente as propriedades imobiliárias. Para os autores, a cadeia de blocos (*blockchain*) não seria suficiente para garantir segurança jurídica e a transmissão de propriedade, sendo necessário recorrer a terceiros, no caso o Estado e a respectiva autoridade judicial/extrajudicial (González; Fernández, 2020, p. 344).

Indo na linha do que defende Werbach (2018), mencionado acima, González e Fernández (2020, p. 422) argumentam que a *blockchain* deve caminhar ao lado do aparato jurídico-legal, compreendida como uma função de apoio na operacionalização de transações imobiliárias, contribuindo para o desenvolvimento do sistema como um todo.

Na atualidade, já existem atividades que são realizadas em conjunto, de forma híbrida, conjugando elementos da inteligência artificial e da atividade notarial e registral. A qualificação híbrida, defendida por Miranda e exposta acima, é a principal delas, buscando-se automatizar e dar eficiência à análise feita na qualificação registral²². Várias

outras atribuições poderão ser repensadas, buscando-se a aprimoração e a modernização do serviço extrajudicial.

Muitas são as dúvidas, mas a conclusão parece ser uníssona: há a necessidade de aprimoramento e discussão da temática, com a qual se presta o artigo. Coaduna-se com o posicionamento de Brandelli, ao defender a coexistência entre a inteligência artificial e a atividade notarial e registral, que manter-se-á pela sua própria natureza intelectual. Destaca-se o seguinte trecho do autor (Brandelli, 2021, p. 398):

Na medida em que a definição das atividades sensíveis, que deverão ficar sob a batuta humana, tende a ser, como se disse, uma escolha social, é importante saber se a sociedade conseguirá em algum momento perceber a importância, a vitalidade, da atividade registral imobiliária para a manutenção e evolução da sociedade economicamente organizada, a fim de que a possa reconhecer como sensível [...] Talvez aí, e não propriamente no avanço tecnológico, resida o principal desafio para os registradores de imóveis: *conseguir tornar óbvio para todos o que o é somente para algumas poucas pessoas.* (grifos nosso)

São inegáveis os avanços trazidos pela tecnologia, mas também os seus respectivos problemas. Tecnologia sem regulação não traz benefícios à sociedade como um todo, devendo-se buscar a conciliação entre a disrupção e o direito, o que é essencial para que ambos evoluam.

Considerações finais

A prática jurídica deve ousar, buscando sempre ser consentânea com a realidade.

Os notários e os registradores devem se adaptar à contemporaneidade, buscando alternativas aos obstáculos vigentes e almejando a efetivação/consecução de direitos. Devem, portanto, pautar a sua atuação pelo interesse público, como própria expressão da função pública que lhes foi delegada.

Todavia, ainda há muito a ser feito. É necessário pensar novas práticas, reinventar-se e ser criativo, trazendo o tema da tecnologia e da inteligência artificial com mais recorrência para a seara acadêmica/doutrinária. Mais importante do que criar condutas, é fazer com que as mudanças sejam efetivadas e, principalmente, estar em constante

adaptação à realidade, com enfoque na implementação permanente da tecnologia à atividade cartorária.

Referências

- ALMEIDA, Carlos Ferreira de. *Publicidade e teoria dos registros*. São Paulo: Almedina, 1966.
- ARRUNÁDA, Benito. *Análisis económico del notariado*. Universidad Pompeu Fabra, Madrid, 1995.
- BARBOSA, Mafalda Miranda et al. (Org.). *Direito Digital e inteligência artificial: diálogos entre Brasil e Europa*. Belo Horizonte: Foco, 2021.
- BONILHA FILHO, Márcio Martins; DUARTE, Andrey Guimarães. *A revolução tecnológica e o direito notarial – o impacto da tecnologia na atividade notarial e a sua regulação pelo Provimento 100 da Corregedoria Nacional de Justiça*. In: ABELHA, André (Org.). *Estudos em Direito Imobiliário – homenagem a Sylvio Capanema de Souza*. São Paulo: Editora Ibradim, 2020, p. 389-400.
- BRANDELLI, Leonardo. *Inteligência artificial e o Registro de Imóveis*. In: GALHARDO, Flaviano et. al. (Org.). *Direito Registral e novas tecnologias*. Rio de Janeiro: Forense, 2021, p. 381-400.
- CAMPILONGO, Celso Fernandes. *Função social do notariado: eficiência, confiança e imparcialidade*. São Paulo: Saraiva, 2014.
- CENEVIVA, Walter. *Novo registro imobiliário brasileiro*. São Paulo: Revista dos Tribunais, 1979.
- CHEZZI, Bernardo (Org.). *Ato eletrônico: notas e registros*. São Paulo: Ibradim, 2021.
- CHEZZI, Bernardo. *Aplicação da LGPD ao Registro de Imóveis*. In: GALHARDO, Flaviano et. al. (Org.). *Direito Registral e novas tecnologias*. Rio de Janeiro: Forense, 2021.
- FARIA, Bianca Castellar de; LACERDA, Naurican Ludovico. *O custo das transações imobiliárias nos Estados Unidos*. *Revista de Direito Imobiliário*. v. 92, p. 59-77, jan.-jul. 2022.
- FÜRST, Maria Eduarda; BÜRGER, Marcelo Luiz Francisco. *Inteligência artificial: conceitos introdutórios e algumas de suas aplicações*. In: EHRHDART JÚNIOR, Marcos; CATALAN, Marcos; NUNES, Cláudia Ribeiro Pereira (Org.). *Inteligência artificial e relações privadas: possibilidades e desafios*. v. 1, Belo Horizonte: Fórum, 2023, p. 19-40.
- GÁLLIGO, Javier Gómez. *Registro de Imóveis e internet*. In: GALHARDO, Flaviano et. al. (Org.). *Direito Registral e novas tecnologias*. Rio de Janeiro: Forense, 2021, p. 249-267.
- GONZÁLEZ, Fernando P. Méndez; FERNÁNDEZ, Luis A Gallego. *Contratos automatizados, cadenas de bloques y registros de la propiedad*. *Revista de Direito Imobiliário*. v. 89, p. 315-430, jul.-dez. 2020.
- GUSTIN, Miracy Barbosa de Sousa; DIAS, Maria Tereza Fonseca; NICÁCIO, Camila Silva. *(Re)pensando a pesquisa jurídica: teoria e prática*. 5. ed. São Paulo: Almedina, 2020.

HAENSEL, Taimi. *A figura dos gatekeepers: aplicação às instituições intermediárias do mercado organizado de valores mobiliários brasileiro*. Dissertação (Mestrado em Direito). Faculdade de Direito da Universidade de São Paulo. São Paulo, 2014.

INTERNATIONAL UNION OF NOTARIES. Website, 2023. Disponível em: <https://www.uinl.org/>. Acesso em: 22 nov. 2023.

JACOMINO, Sérgio; UNGER, Adriana Jacoto. *NFT's – a tokenização imobiliária e o metaverso digital*. Disponível em: https://www.migalhas.com.br/arquivos/2022/4/7CEF9CFF30BB7E_Tokenizacao.pdf. Acesso em: 10 ago. 2023.

KERN, Marinho Dembinski; COSTA JUNIOR, Francisco José de Almeida Prado Ferraz. *Princípios do Registro de Imóveis brasileiro*. 2. ed., São Paulo: Revista dos Tribunais, 2022.

KINDEL, Augusto Lermen. *Responsabilidade civil dos notários e registradores*. Porto Alegre: Norton Editor, 2007.

KUMPEL, Vitor et al. *Tratado notarial e registral*. v. 3, São Paulo: YK Editora, 2017.

LAGO, Ivan Jacopetti do. *Antevésperas dos registros imobiliários – um passeio histórico às fontes romanoegípcias*. In: CANDAU, Alfonso et al. (Org.). *Estudos em homenagem a Sérgio Jacomino*. GestLegal: Lisboa, 2022, p. 173-194.

LOPES, Miguel Maria de Serpa. *Tratado dos registros públicos: em comentário ao Decreto n. 4.857, de 9 de novembro de 1939, com as alterações introduzidas pelo Decreto n. 5.318, de 29 de novembro de 1940 e legislação posterior em conexão com o Direito Privado brasileiro*. Brasília: Brasília Jurídica, 1997.

LOUREIRO, Luiz Guilherme. *Registros Públicos: teoria e prática*. 10. ed. Salvador: Juspodivm, 2019.

MAZITELI NETO, Celso; BRANDELLI, Leonardo. *Blockchain e o Registro de Imóveis*. *Revista de Direito Imobiliário*. v. 87, p. 63-95, jul.-dez. 2019.

MIRANDA, Caleb Matheus Ribeiro de. *A computação cognitiva e o Registro de Imóveis*. *Revista de Direito Imobiliário*. v. 83, p. 79-100, jul.-dez. 2017.

NALINI, José Renato. *Conclusão: o futuro do Registro de Imóveis*. In: GALHARDO, Flaviano et. al. (Org.). *Direito Registral e novas tecnologias*. Rio de Janeiro: Forense, 2021, p. 447-476.

NUNES, Dierle; LUCON, Paulo Henrique dos Santos; WOLKART, Erik Navarro (Org.). *Inteligência artificial e Direito Processual: os impactos da virada tecnológica no Direito Processual*. 2. ed., Salvador: Juspodivm, 2020.

PATRÃO, Afonso. *E-Justice e transações imobiliárias internacionais na União Europeia*. In: CANDAU, Alfonso et al. (Org.). *Estudos em homenagem a Sérgio Jacomino*. GestLegal: Lisboa, 2022, p. 23-43.

RIBEIRO, Luís Paulo Aliende. *Regulação da função pública notarial e de registro*. São Paulo: Saraiva, 2009.

RICHTER, Luiz Egon; PHILIPPI, Juliana Horn Machado. *Inovação e segurança jurídica: a necessidade de regulação de criptoativos imobiliários e do uso da blockchain nas transações imobiliárias*. *Revista de Direito Imobiliário*. v. 93, p. 191-228, jul.-dez. 2022.

RODRIGUES, Daniel Lago. *Os smart contracts e seu caráter trustless*. In: GALHARDO, Flaviano et. al. (Org.). *Direito Registral e novas tecnologias*. Rio de Janeiro: Forense, 2021.

RODRIGUES, Marcelo. *Tratado de Registros Públicos e Direito Notarial*. 4. ed. São Paulo: Juspodivm, 2022.

SALLES, Marcos Huet Nioac de. *A reinvenção do papel do cartório de imóveis na era da tecnologia blockchain: uma investigação exploratória*. Dissertação de Mestrado – Administração Pública na Fundação Getúlio Vargas. Rio de Janeiro, 2019.

SCHWAB, Klaus. *A quarta revolução industrial*. São Paulo: Edipro, 2016.

SHILLER, Robert. *The subprime solution: how today's global financial crisis happened, and what to do about it*. Oxford: Princeton University Press, 2008.

TUTIKIAN, Cláudia Fonseca. *Propriedade imobiliária e o registro de imóveis: perspectiva histórica, econômica, social e jurídica*. São Paulo: Quartier Latin, 2011.

UINL. *Fundamental Principles*. Disponível em: www.uinl.org/principio-fundamentales. Acesso em: 22 nov. 2023.

VALE, Luís Manoel Borges do; MAZZOLA, Marcelo. Os novos paradigmas da execução fiscal, na era da inteligência artificial. In: BELLIZZE, Marco Aurélio et al. (Org.). *Execução Civil: novas tendências*. Belo Horizonte: Foco, 2022, p. 447-460.

VALE, Luís Manoel Borges do; PEREIRA, João Sergio dos Santos Soares. *Teoria geral do processo tecnológico*. São Paulo: Revista dos Tribunais, 2023.

WERBACH, Kevin. *Trust, but verify: why the blockchain needs the law*. *Berkely Technology Law Journal*. v. 33, p. 489-552, 2018.

SALLES, Marcos Huet Nioac de. *A reinvenção do papel do cartório de imóveis na era da tecnologia blockchain: uma investigação exploratória*. Dissertação (Mestrado em Administração Pública). Rio de Janeiro: Fundação Getúlio Vargas, 2019.



■ Giselle Beiguelman, frames do vídeo generativo da série Flora Rebellis, versão 2024.
Imagens geradas com Inteligência Artificial (Style Gans)

■ [Acesse o vídeo clicando na imagem](#)





IMAGINAÇÃO E INTELIGÊNCIA:

Apropriação, hibridismos, realidades e Inteligência Artificial em uma exposição artística

DOI: <https://doi.org/10.35699/2965-6931.2023.47704>

PABLO GOBIRA*; **ANA LUIZA PEDROSA CAMILO****; **EDUARDO SÉLLOS*****

RESUMO: Este artigo objetiva apresentar a exposição AI.MAGINATION, aberta no dia 10 de abril de 2023. A exposição foi realizada pelo LabFront, um grupo de pesquisa, desenvolvimento e inovação que atua nas fronteiras entre arte, ciência e tecnologia. A AI.MAGINATION é uma exposição digital com um recorte curatorial que explora o uso da Inteligência Artificial (IA) no campo das artes digitais. O conceito da exposição foi desenvolvido na interseção entre disciplinas ofertadas, durante o segundo semestre de 2022, na graduação em artes na Escola Guignard (UEMG), resultantes de pesquisas realizadas. A exposição é composta por obras de artistas iniciantes e consolidados com atuação nacional/internacional e suas produções poéticas que utilizam IA. A curadoria também é uma experiência construída com IA e projetada em um ambiente online, 3D, conectado a redes blockchain. Dessa forma, este artigo apresenta uma experiência de uso generalizado de Inteligência Artificial no campo da arte.

PALAVRAS-CHAVE: artes digitais; inteligência artificial; curadoria; exposição de arte; metaverso.

Imagination and intelligence: Appropriation, hybridisms, realities and Artificial Intelligence in an artistic exhibition

ABSTRACT

This article aims to present the AI.MAGINATION exhibition, which opened on April 10, 2023. The exhibition was held by LabFront, a research, development, and innovation group that operates on the borders between art, science, and technology. AI.MAGINATION is a curatorial digital exhibition that explores the use of Artificial Intelligence (AI) in the field of digital arts. The exhibition concept was developed at the intersection between disciplines offered, during the second half of 2022, in the graduation in arts at Escola Guignard (UEMG), resulting from research. The exhibition comprises works by beginning and established artists with national/international performance and their poetic productions that use AI. Curation is also an experience built with AI and designed in an online, 3D environment connected to blockchain networks. Thus, this article presents an understanding of the widespread use of Artificial Intelligence in art.

KEYWORDS: digital arts; artificial intelligence; curation; art exhibition; metaverse.

* Universidade Federal de Minas Gerais / ** Universidade Federal de Minas Gerais / *** Centro Universitário Una.

I Neste trabalho não exploraremos a dimensão educacional (STEAM) aplicada no processo de desenvolvimento das disciplinas relacionadas a exposição AI.MAGINATION porque o fizemos em outro trabalho, além de extrapolar o nosso objetivo aqui de focar o uso das IA na exposição. O artigo será publicado como capítulo de livro que está no prelo (GOBIRA, CAMILO, s/d). O livro, que enfoca em seus capítulos vários aspectos e aplicações das IA na educação, tem previsão de ser publicado pela editora da Universidade Federal da Bahia ainda em 2023. Para interessados, sugerimos acompanhar as redes sociais do LabFront e o nosso website: <https://labfront.weebly.com/>

Introdução

O Laboratório de Poéticas Fronteiriças (LabFront - UEMG/CNPq) é um grupo de pesquisa, desenvolvimento e inovação que se propõe a problematizar as/nas fronteiras e atua na consolidação de uma área de pesquisa, produção artística e teórica dos saberes, das linguagens artísticas, do conhecimento científico e tecnológico. A exposição AI.MAGINATION, produzida pelo LabFront, sob curadoria e coordenação de Pablo Gobira, faz parte do Multilab, um projeto extensionista da Escola Guignard (UEMG) que busca expandir as possibilidades de diálogos entre o que é realizado na universidade e fora dela abrangendo o contexto das artes.

O objetivo deste artigo é apresentar a exibição AI.MAGINATION, aberta em 10 de abril de 2023 como uma exposição de longa duração. AI.MAGINATION tem a participação de artistas iniciantes e veteranos conhecidos no campo das artes digitais com recepção nacional e internacional. Ela é uma exposição com obras desenvolvidas em relação com Inteligência Artificial (IA) e tem a sua curadoria pensada a partir das potencialidades das IA. A exposição AI.MAGINATION é uma proposta artística e cultural que surge a partir das disciplinas “Ateliê Transartes: hibridismos e apropriações artísticas em realidades diversas” e “Curadoria em arte digital” ministradas por Pablo Gobira e cujo conteúdo surge de projetos de pesquisa coordenado por este no Laboratório de Poéticas Fronteiriças.

Há vários caminhos para pensar a relação da arte com IA. Aqui decidimos abordar essa exposição e os usos da IA na curadoria e arte, mas destacamos que a proposta dialoga com a ação de ensino de arte e IA em uma perspectiva STEAM (*science, technology, engineering, arts and mathematics*), para estudantes ainda não imersos nessa integração de áreas em seus processos educacionais formativos¹. Assim, apresentaremos aqui como a exposição explora diferentes formas de utilização da Inteligência Artificial no campo das artes digitais e reflete sobre a apropriação, o hibridismo e a construção de realidades por meio das ferramentas inteligentes.

De acordo com Gobira (2022a, p. 47) “o termo artes digitais refere-se a uma área na qual a obra de arte é criada a partir da sua relação com a ciência e tecnologia após o surgimento do computador no século XX”. Em seu contexto mais abrangente, a arte

digital também pode compreender produções como “webarte, netarte, ciberarte, bioarte, gamearte, instalações interativas, mídias locativas e outras atividades relacionadas” (Ministério da Cultura, 2009, s/p *apud* Gobira, 2022a, p. 47).

Os conceitos “apropriação”, “hibridismos” e “realidades” compuseram não apenas a AI.MAGINATION, mas fundamentaram o percurso metodológico dos estudantes durante as aulas. Com base nas discussões sobre teoria e história da arte foi possibilitado aos estudantes constituir repertório teórico e prático sobre as experimentações poéticas entre ciência e tecnologia até a arte contemporânea.

No século XX, com as vanguardas e o impulsionamento das técnicas de reprodução, a apropriação como forma de criticar o sistema da arte se tornou mais latente através de experimentações poéticas que propunham novas realidades que escapavam, muitas das vezes, do sistema da arte. Dessa forma, a AI.MAGINATION foi construída com base nessas reflexões e conceitos teóricos/práticos trabalhados nas disciplinas que foram potencializados por intermédio do uso colaborativo de IA.

Assim, para garantir a coerência conceitual da exposição, as obras expostas e todo o processo de criação (identidade visual, design de ambientes e expografia) tiveram em seus processos a operação de ferramentas inteligentes como instrumento mediador criativo. Esse caminho curatorial experimentado nas disciplinas e na exposição permitiu construir a discussão sobre as ferramentas inteligentes com maior profundidade, levantando reflexões, a produção poética e as possíveis recepções no sistema da arte.

De acordo com Bulhões (1995, p. 114), o sistema da arte alinha-se a um conjunto de pessoas e organizações responsáveis pela criação, difusão e consumo de objetos e eventos. Entendendo isso, é importante refletir sobre os grupos de indivíduos que definem o que é arte, levando em consideração que não existe uma única definição sobre seu conceito. De toda forma, esses grupos estabelecem “critérios e valores da arte para toda uma sociedade ao longo de determinado período” (Bulhões, 1995, p. 114).

As ferramentas inteligentes utilizadas neste trabalho são *softwares* abertos e/ou privados que resultaram em diferentes possibilidades criativas de acordo com suas operações artísticas. Dentre diversos sistemas, se fazem presentes o uso do *ChatGPT*, *Dall-e*, *Dall-e 2*, *Dall-e mini*, *Google Colab*, *Hugging Face*, *Midjourney*, *Notevibes*, *OCR*, *Stable Diffusion*, *TikTok voice generator* e *Wonder*. Ao longo deste trabalho a apropriação dessas ferramentas serão melhor apontadas, mas ainda assim, julgamos importante

oferecer descrições curtas a respeito das ferramentas para compreensão as suas possibilidades técnicas no campo das artes digitais.

O *ChatGPT (text-to-text)* propicia conversas criativas que exploram narrativas que se baseiam em sua *Big Data* alimentada com informações encontradas na internet até 2021. A ferramenta *Dall-e*, e suas variantes, são capazes de gerar imagens únicas (generativas) a partir de *prompts* de comando. Esse formato de sistema inteligente é compreendido como *text-to-image*. O OCR (*Optical Character Recognition*) tem a possibilidade de reconhecer caracteres a partir de arquivos de imagem. O *Google Colab* facilita a criação de propostas artísticas e permite ao usuário escrever códigos e testar algoritmos. A ferramenta propõe ser um ambiente acessível e oferece a possibilidade do usuário compartilhar sua criação com a comunidade. A plataforma *Hugging Face* se relaciona com a experimentação de modelos de linguagem pré-treinados. Assim, as informações disponibilizadas na plataforma podem ser apropriadas, personalizadas e adaptadas para necessidades artísticas. As ferramentas *Midjourney* e *Wonder* são sistemas com possibilidades *text-to-image* que podem auxiliar em criações de narrativas visuais. O *Stable Diffusion* é uma ferramenta inteligente escrita em Python que pode gerar imagens ou provocar transformações visuais. A dimensão sonora pode ser experimentada através do *TikTok Voice Generator* e *Notevibes*, sendo ambas ferramentas de síntese de voz baseada em IA (*text-to-audio*). É importante destacar que, em alguns processos criativos que envolvem a exposição AI.MAGINATION, os artistas e a equipe curatorial se apropriam de mais de uma ferramenta inteligente, instituindo assim uma ampla rede colaborativa.

Para alcançarmos o objetivo de abordar a exposição AI.MAGINATION, o artigo está organizado em cinco seções, sendo a primeira esta introdução. A segunda seção deste trabalho trata sobre o *status quo* da curadoria e as insurgências latentes desta área. Subsequente a isso, a terceira seção expressa os processos de criação da exposição AI.MAGINATION, contemplando também as obras, os artistas envolvidos e maiores levantamentos sobre o uso de IA no projeto. A quarta seção é dedicada às considerações finais e na última são encontradas as referências que alicerçam o artigo.

a curadoria nas artes digitais

O desenvolvimento da curadoria e a arte

De acordo com Bettina Rupp (2010), a curadoria, em sua origem, dispõe de um caráter organizacional responsável por agrupar as “curiosidades” que abarrotavam os gabinetes dos colecionadores com seus artefatos arrematados de diferentes partes do mundo. A organização das coleções poderiam partir de diferentes aspectos, ainda que de maneira limitada pela região de onde advinham ou pela área de conhecimento com a qual se relacionavam (Rupp, 2010, p. 9). Ainda de acordo com a autora, esse comportamento classificatório do curador permaneceu com o surgimento dos museus, em função do objetivo de conservação das obras que compunham o acervo das instituições (Rupp, 2010). Porém, julgamos importante destacar que esse comportamento classificatório era conflitante com as perspectivas dos artistas, tendo em vista que a ideia de curadoria pareceria se distanciar de uma visão cotidiana de arte se aliando a uma dimensão institucional.

A virada do século XIX para o XX veio acompanhada das iniciativas vanguardistas, cuja disrupção para com o sistema da arte vigente se deu por diversas experimentações que incluem a apropriação de múltiplas técnicas, métodos/meios, materiais etc.. Um fator para essas iniciativas será a emancipação do uso ritualístico da arte, oferecendo maior disponibilidade do que compõe o mundo através da reprodutibilidade técnica (Benjamin, 1987, p. 173).

Como observado por Benjamin (1987, p. 166), “em sua essência, a obra de arte sempre foi reprodutível. O que os homens faziam sempre podia ser imitado por outros homens”. Para dessacralizar tal sistema, alguns dos exercícios praticados por grupos vanguardistas se baseiam no apoderamento conveniente de mercadorias para criticar as forças legitimadoras do sistema da arte (Debord; Wolman, 1956, p. 1-6). Assim, por meio desses experimentos, os aspectos museológicos da época foram fortemente confrontados (Marmo; Lamas, 2012) por artistas em suas perspectivas vanguardistas.

Bettina Rupp (2011) conta que a partir do caráter criativo percebido na atuação do curador, ele também vai sendo reconhecido como um artista. Nesse sentido, compreendemos que as experimentações vanguardistas do século XX tiveram influência

no papel do curador e na sua relação com a arte. Um exemplo marcante trazido pela autora sobre esse fenômeno é a exposição *First papers of Surrealism*, curada por Marcel Duchamp em 1942, em Nova York. O artista incorpora a sua própria obra, intitulada *His twine*, à exposição realizando uma interferência tão marcante que “a mostra acabou caracterizada pelo emaranhado de fios brancos que Duchamp amarrou pelos expositores onde estavam as pinturas, dificultando a circulação na galeria e também a visibilidades dos quadros.” (O’doherly *apud* Rupp, 2011, p. 135). Com este exemplo, Rupp (2011, p. 5) mostra que, nessa experiência, o público se via imerso na escolha curatorial para apresentar a obra, demonstrando a possibilidade do curador em transcender o ato de organizar a exposição para construir uma vivência mais ativa.

Apropriação, hibridismo e criação de realidades

Na exposição AI.MAGINATION, o recorte curatorial abraçou dimensões encontradas na teoria e história da arte, sobretudo no último século: a apropriação, os hibridismos e a criação de novas realidades.

A apropriação e o hibridismo são conceitos caros para a teoria e história da arte. São nessas contingências que compreendemos a importância das críticas artísticas para com a realidade vigente, pois as vanguardas também nos ensinam como construir de forma prática múltiplas realidades que tensionam o cotidiano em seus aspectos políticos, econômicos e sociais (Gobira, 2022b, p. 180).

A importante teórica Lucia Santaella (2008, p. 20), no contexto de avanços tecnológicos, aborda o hibridismo relacionado à interconexão dos espaços físicos com os espaços virtuais, caracterizando uma ecologia pluralista das mídias locativas. Segundo a autora, o conceito de “mídias locativas” é definido pela “Mdiclopedia (sic) digital das novas mídias (...) como tecnologias baseadas em lugares, ou seja, tecnologias sem fio, tecnologias de vigilância, de rastreamento e de posicionamento que permitem que a informação seja ligada a espaços geográficos” (Santaella, 2008, p. 22). Dessa forma, as mídias locativas podem ser concepções que permeiam e conceituam possibilidades expositivas de obras de artes digitais. Santaella (2008, p. 20) aponta os termos “híbrido”,

“hibridismo”, “hibridação” e “hibridização” em relação com as múltiplas linguagens na “hipermídia”, principalmente depois da explosão das redes de comunicação e da *World Wide Web*.

O advento da internet intensificou a relação da arte e tecnologia exponencialmente, pois a produção e disponibilidade de informação acelera e impacta todas as áreas de conhecimento na sociedade, tornando a influência do digital inevitável (Santaella, 2018). Nesse cenário, Lucia Santaella acredita que “a arte digital, no contexto intrincado da arte contemporânea, enroscava-se e enrosca-se crescentemente em um hibridismo de tal modo radical a ponto de se tornar inextricável” (2021, p. 62). Nesse sentido, o conceito de ciberespaço é fruto dessas noções híbridas na hipermídia e “consiste de uma realidade multidirecional, artificial ou virtual incorporada a uma rede global, sustentada por computadores que funcionam como meios de geração e acesso” (Santaella, 2008, p. 21).

Pablo Gobira evidencia que no cenário atual o espaço se configura “a partir do entrelaçamento entre realidade e virtualidade, sendo esta entendida também como parte do mesmo espaço geofísico” (2018, p. 87-88). Esta perspectiva destaca a natureza híbrida existente no campo das artes digitais e em outras áreas que se articulam com esse campo. No sentido poético, a infinidade de possibilidades advindas do entrelaçamento entre arte, ciência e tecnologia seguem tendo distintas recepções, pois outras perspectivas poéticas surgem e confabulam múltiplas compreensões acerca da arte e na prospecção de novas realidades. Ao tratarmos de realidades, é importante entender que há muitos modos de conceituar o que é realidade. De acordo com Peter Berger e Thomas Luckmann (1985, p. 87) “a sociedade é um produto humano. A sociedade é uma realidade objetiva. O homem é um produto social”. Com este trecho curto já percebemos que para os autores a realidade é uma construção social.

A criação de uma realidade diferente do nosso cotidiano é uma pauta do cenário artístico ao menos desde o século XIX. No século XX, o diálogo do artista com sua realidade se intensifica através de “incursões significativamente inspiradas em acontecimentos, tal como Pablo Picasso em sua célebre obra *Guernica* (1937)” (Nunes, 2018, p.56). Já no século XXI, na perspectiva incorporada ao sistema da arte, portanto, dentro de uma lógica da economia política vigente, não existem dúvidas de que as tecnologias digitais foram capazes de perfurar, sobrepor ou refazer a realidade social (por exem-

2 Neste trabalho, por necessidade de recorte do nosso relato, não abordaremos a ideia de metaverso. Para caso haja interesse em aprofundar nessa reflexão ver “Avatarização da vida, biografia, impurezas e metaverso”, de Pablo Gobira (2023).

plo, com a realidade virtual imersiva, a realidade aumentada, a realidade mista, ou na ideia de metaverso² etc.).

A realidade aumentada, por exemplo, de acordo com Juan Aristizábal (2021, p. 154), é definida como:

Uma visualização do mundo real com uma variação virtual do ambiente. Esta tecnologia permite que o usuário possa observar o mundo real com objetos virtuais sobrepostos ou na composição do ambiente, é produzida uma coexistência ou combinação entre objetos virtuais e reais que são combinados por meio de ferramentas informáticas para que possamos obter uma realidade mista em tempo real (MAPLE, 2014 apud ARISTIZÁBAL, 2021, p. 154).

Sendo assim, a “sobreposição” de realidades é um exemplo concreto do hibridismo no campo das artes digitais, em que diferentes linguagens e mídias se intrincam e criam novas expressões artísticas. Aristizábal (2021, p. 155) afirma ainda que o avanço tecnológico segue expandindo as percepções dos seres humanos, pois existe o interesse em melhorar e criar novas capacidades de experiência da realidade e ampliação da própria corporeidade. Já Santaella (2021, p. 66) destaca a “automatização e expansão de capacidades cognitivas humanas por meio de tecnologias de aprendizagem de máquina (*machine learning*) e computação cognitiva”.

A autora, no texto “Arte, ciência e tecnologia: um campo em expansão” (Santaella, 2018), observa como o desenvolvimento do digital tem uma influência direta na atualização da linguagem, implicando a alteração de toda a constituição humana. Para Santaella (2018), a eclosão da *mass media* e o constante surgimento de novos recursos tecnológicos impulsionou artistas a explorar aparelhos digitais (televisores, rádio, entre outros) como suporte para produzir arte. É nesse cenário interdisciplinar que surge o que é entendido como arte digital, um campo que relaciona arte, ciência e tecnologia para a criação de novas poéticas (Santaella, 2018).

Na exposição AI.MAGINATION, os conceitos de apropriação, hibridismo e realidades puderam ser incorporados, levando em conta o percurso metodológico das disciplinas de ateliê transartes e curadoria em arte digital. Assim, podemos ver as ligações desses conceitos com o campo das artes digitais onde, de acordo com Gobira (2022a, p. 47-48), “uma obra pode ser criada por artistas, engenheiros, cientistas - em cooperação. Podendo envolver Inteligência Artificial, sistemas inorgânicos ou orgâni-

cos alterados através de engenharia genética”. Esses usos, ainda nas palavras de Gobira (2022a, p. 48), são “capazes de constituir formas de vida artificial expressas em uma instalação, em uma performance ou de outros modos”.

Aplicando esse entendimento nas aulas a equipe técnica da curadoria faz com que a AI.MAGINATION explore a apropriação de elementos de diferentes contextos artísticos e tecnológicos, usando essas abordagens como ponto de partida para a criação de expressões híbridas. Para a exposição foram combinadas diferentes linguagens artísticas, tecnologias e mídias com o uso de IA. Essa abordagem, que funde a criatividade humana e a capacidade das máquinas (*machine learning* e *deep learning*), oferece ao público possibilidades de ampliação das capacidades perceptivas e cognitivas, uma vez que, por meio da exposição, é possível criar, prospectar e refletir sobre diversas realidades. Lucia Santaella (2021, p. 66) define IA como “uma área da ciência computacional que leva as máquinas a executar tarefas similares àquelas desempenhadas pela inteligência humana, tais como percepção visual, tomada de decisão, tradução, reconhecimento de voz etc.”.

Em síntese, a exposição, ao ser constituída por experimentações do campo de artes digitais, tem como objetivo dialogar com aspectos híbridos para conceber novas realidades ao se apropriar das IA. Ainda que este artigo não se aprofunde nas perspectivas preservacionistas de obras de artes digitais expostas, compreendemos que cabe à atuação curatorial colaborar com este ponto. Afinal, com a exposição percebe-se a pluralidade de expressões poéticas que o recorte curatorial abarca. Por isso, para manter a consistência da exibição aqui demarcada, foram estudados diferentes usos de ferramentas inteligentes para que elas pudessem ser treinadas de forma a atender as expectativas da AI.MAGINATION. Logo, essas informações, advindas do processo de criação da exposição, permitem ao curador estabelecer parâmetros de preservação digital de obras de artes digitais, por exemplo, no caso da sua conservação em acervo digital.

Ai.magination

3 Participaram de modo direto da equipe técnica curatorial: Pablo Gobira, Ana Luiza Pedrosa Camilo, Vinícius Viana Rafael, Eduardo Séllos e Gabriel Rios. Porém, essa equipe teve o apoio dos demais membros do Laboratório de Poéticas Fronteiriças, sobretudo Emanuelle de Oliveira Silva e Priscila Rezende Portugal, em nome das quais agradecemos a todos os membros do grupo pelo apoio.

4 Como o nosso objetivo é apresentar a relação da arte com a Inteligência Artificial em uma exposição, não abordaremos a infraestrutura tecnológica do ambiente 3D utilizado neste artigo. Infelizmente, teremos que deixar para realizar essa discussão em uma próxima oportunidade.

O trabalho para a criação da exposição “AI.MAGINATION: apropriação, hibridismo e realidades” (<https://exposicaoaimagination.weebly.com/>) teve início em agosto de 2022, e atravessou as aulas das disciplinas de “Ateliê transartes: hibridismos e apropriações artísticas em realidades diversas” e “Curadoria em arte digital”. Para a sua criação, trabalhamos com uma equipe técnica composta por membros do Laboratório de Poéticas Fronteiriças⁴ que também cursaram as disciplinas. Em formato digital, tridimensional através da plataforma online *Voxels*, em redes *blockchain*⁵, a AI.MAGINATION tem como proposta ser uma exibição de longa duração.

A exposição em formato digital permite o acesso do público tanto por dispositivos móveis quanto estáticos, com uso de estereoscopia ou de tela tradicional. A equipe técnica da AI.MAGINATION se reuniu para discutir o projeto curatorial e o modo como poderiam incorporar as IA nesse projeto, auxiliando a produção da exposição, sua identidade visual, o ambiente construído digitalmente etc. Optou-se por utilizar as IA: *ChatGPT*, *MidJourney* e *Stable Diffusion* para essa produção. Essas IA foram aplicadas de diferentes maneiras na produção da identidade visual, do espaço expográfico, da curadoria, da divulgação e das obras, com o objetivo de demonstrar o potencial desses instrumentos infocognotecnológicos.

Através dos estudos da operação dessas IA, de como estavam sendo usadas e de como seriam melhor utilizadas em uma proposta curatorial, as aplicamos no intuito de explorar as interseções entre arte e IA, tornando acessível os debates e conceitos da nossa sociedade nessas relações.

3.1 PRODUÇÃO CURATORIAL E IA

Os artistas em exibição também utilizaram uma variedade de técnicas, processos colaborativos e ferramentas para criar os trabalhos apresentados. Em um projeto como esse, o ambiente tridimensional construído digitalmente acaba sendo um elemento essencial da seção de expografia do projeto curatorial. A partir da pesquisa e experimentação de como as IA poderiam auxiliar na criação dos ambientes, a equipe agrupou re-

ferências retiradas das propostas geradas na *Midjourney*, constituindo uma expografia dividida em três andares de um edifício (Figuras 1, 2 e 3) e em um ambiente externo (Figuras 4, 5, 6 e 7).

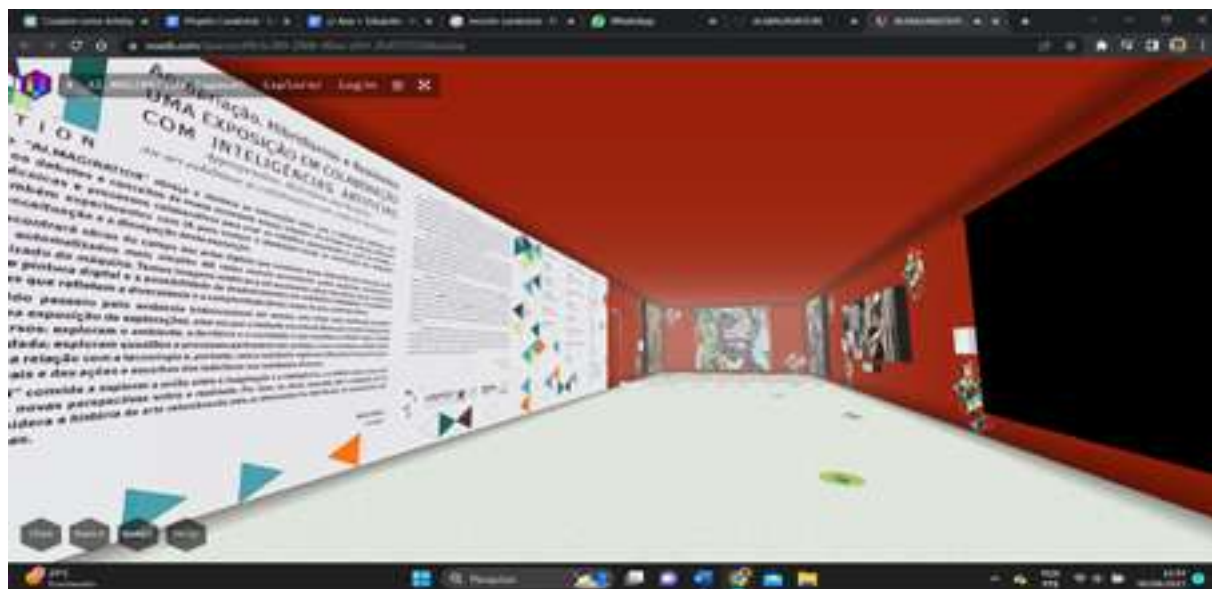


Figura 1 - Vista do primeiro andar da exposição AI.MAGINATION. Obras: Mitômato, de Fabio FON (2023); Interagindo com objeto mal identificado, de Dera (2023); Sem título, de Gustavo Machado (2023). Fonte: Arquivo dos autores.



Figura 2 - Vista do segundo andar da exposição AI.MAGINATION. Obras: Procura-se Helene Alberti, de Larissa Campello (2023); Manacá-de-cheiro, de Gilberto Prado e Suzete Venturrelli (2021); I.A. sem fio, de Eduardo Séllos (2023); Série Cartas do inferno, de Chico Marinho (2023). Fonte: Arquivo dos autores.



Figura 3 - Vista do terceiro andar da exposição AI.MAGINATION. Obras: Mosaico e releitura da obra Stánczyk de Jan Matejko, de Skrif (2023); Consciência Alada, de Ana Luiza Pedrosa Camilo (2023). Fonte: Arquivo dos autores.



Figura 4 - Vista da área externa da exposição AI.MAGINATION. Obras: Brasília em um mundo pós-apocalíptico, de José Dias (2023); Até agora, a violência, a guerra, a pilhagem, o assassinato, o roubo, etc, foram aceitos como escolha motriz da história, de Luiz Carlos (2021); Rua lá perto da casa da sua avó, de Gabriella Pedro (2023). Fonte: Arquivo dos autores.



Figura 5 - Vista da área externa da exposição AI.MAGINATION. Obras: Iconografia da construdestruição, de Vinicius “Apos” Viana (2023); Nova velha vida I, de Gabriel Rios (2023). Fonte: Arquivo dos autores.



Figura 6 - Vista externa e aérea das obras Brasília em um mundo pós-apocalíptico (2023), de José Dias e Rua lá perto da casa da sua avó (2023), de Gabriella Pedro Fonte: Arquivo dos autores.



Figura 7 - Vista aérea da parte externa da galeria Fonte: Arquivo dos autores.

Para dar forma ao espaço expositivo, ferramentas inteligentes como o *ChatGPT* (na época ainda *Generative Pre-Training Transformer 3 - GPT-3*) foram utilizadas experimentando e criando através dele possíveis *prompts* de comando. A partir desses testes, por meio da plataforma Discord, onde se opera o *Midjourney*, foram aplicados *prompts* ligados a criação de realidades e a experiência imersiva em museus e galerias, buscando agrupar imagens que permitissem vermos como essa, que é uma das plataformas “*text-to-image*” mais utilizadas, apresentava propostas nesse sentido. Para que essas imagens geradas pela IA pudessem ser incorporadas ao ambiente 3D do *Voxels*, foi necessário adaptar a expografia devido ao formato de “*pixel tridimensional*” (*voxel*) dessa plataforma.

Para além da apropriação e experimentos com os elementos sugeridos pela IA, foi necessário realizar outras escolhas curatoriais devido às especificidades das obras expostas. Portanto, foram escolhidas as proporções das galerias, a divisão entre galeria externa e interna, a quantidade de andares, as formas das construções em geral foram determinadas com base nas necessidades apresentadas pelo curador diante das obras e do recorte curatorial realizado (Figura 7).

Outro elemento tratado a partir das IA foi a identidade visual. Neste caso, usou-se uma base de imagens no *Stable Diffusion*⁵ relacionado às próprias obras dos artistas

para se gerar paleta de cores e elementos visuais para compor a identidade visual. O projeto gráfico, portanto, segue essa identidade e dialoga profundamente com as obras expostas. Assim, as cores utilizadas, tipografia e grafismos da exposição surgem de um processo de treinamento de máquina cujos resultados foram submetidos a escolhas e tratamentos técnicos feitos pela equipe que trabalhou na comunicação visual da exposição. Os grafismos gerados, por exemplo, foram aplicados na expografia com objetivo de auxiliar o público a percorrer o caminho expográfico sugerido, assim, os elementos gráficos escolhidos contribuem para a ambientação e experiência no espaço⁶.

No texto do curador da exposição, disponível na galeria do primeiro andar, vemos o seguinte:

A exposição é uma oportunidade única para os visitantes conhecerem obras de gêneros diversos, incluindo desenhos, poesia, instalações, esculturas, fotografias e videoarte, que exploram a concepção de novas realidades, hibridismo e apropriações no campo das Artes Digitais. Para dar forma à exposição AI.MAGINATION, foram utilizadas ferramentas de Inteligência Artificial (IA), incluindo a IA text to image Mid-journey e o Chat GPT, que proporcionaram a criação de diversas possibilidades de exposições, galerias, museus e projetos arquitetônicos.

Com base nas escolhas curatoriais e nas características das obras selecionadas, a exposição AI.MAGINATION foi projetada em um ambiente tridimensional no metaverso Voxels em rede blockchain. A estrutura da galeria é composta por espaços expositivos modulares, galerias internas e externas e andares, criando uma experiência imersiva para os visitantes que a podem experimentar através do navegador nos computadores, smartphones e também através de óculos de estereoscopia. As cores e grafismos aplicados nas galerias e plotagens dialogam com a identidade visual da exposição e contribuem para a ambientação e experiência do visitante.

A exposição AI.MAGINATION é uma oportunidade imperdível para público geral e artistas com formações diversas que tiverem interesse em se aproximar dos debates atuais do campo das Artes Digitais e Inteligência Artificial. Essa exposição reúne e acrescenta obras experimentais que desafiam as fronteiras do que é possível criar e imaginar com as ferramentas inteligentes.

Até aqui observamos as tecnologias digitais representarem a quintessência da exposição, bem como da nossa sociedade contemporânea. Sobre isso, Nunes (2018, p. 70) afirma que os “artistas se apropriam dos significados intrínsecos destes meios para

5 *Stable Diffusion* é um ambiente colaborativo que funciona no navegador através da nuvem, permitindo programar, compartilhar e rodar códigos principalmente na linguagem *Python*. O código do *Stable Diffusion* pode ser utilizado de muitas maneiras, porém o *Google Colab* foi o escolhido neste projeto.

6 Um artigo sobre a construção do ambiente 3D, a identidade visual e o projeto expográfico está sendo escrito e, por isso, não será aprofundado neste texto, além de considerarmos que o relato aqui realizado é suficiente para demarcar os usos de IA nessas produções.

criar suas incursões ficcionais como reflexos de nosso tempo, da maneira como consumimos, das implicações da tecnologia em nosso cotidiano”. Essa afirmação se relaciona com a AI.MAGINATION se considerarmos que ela busca não apenas experimentar a própria visão do mundo contemporâneo, mas também a desafiar as possibilidades intrincadas nas interseções das artes, ciências e tecnologias.

As obras em exibição

Manacá-de-cheiro (2021), é um trabalho realizado pelos artistas Gilberto Prado e Suzete Venturelli. Uma obra de arte generativa em vídeo realizada através de um algoritmo que considera o posicionamento do Sol e reproduz o florescimento da árvore. As flores digitais do manacá-de-cheiro se alternam com imagens da própria árvore. A narração que acompanha o vídeo se relaciona com o aparecimento e o desaparecimento (Figuras 8 e 9).

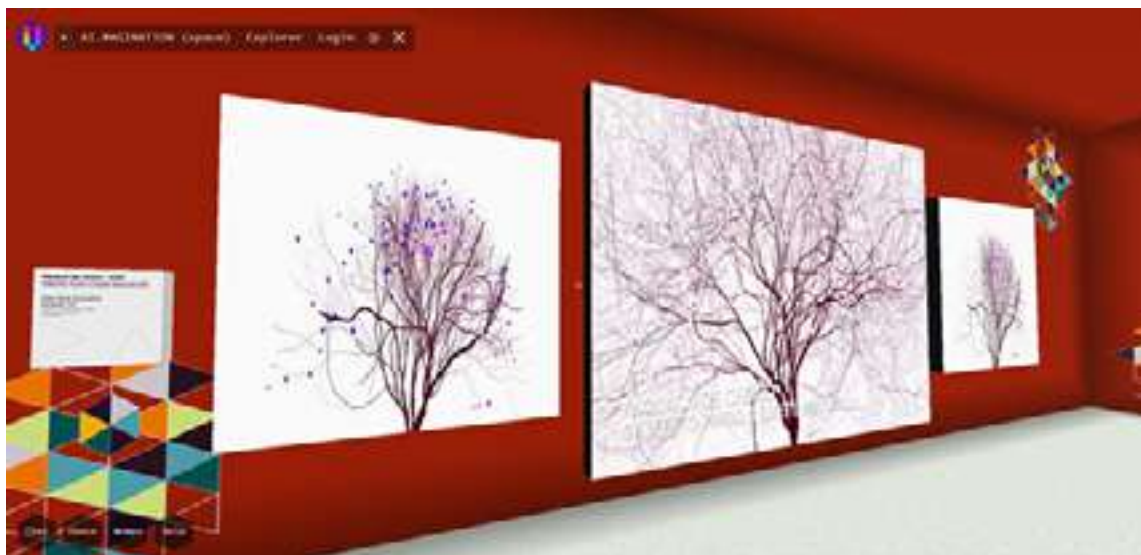


Figura 8 - Vista lateral da obra Manacá-de-cheiro (2021), de Gilberto Prado e Suzete Venturelli

Fonte: Arquivo dos autores.



Figura 9 - Obra Manacá-de-cheiro (2021), de Gilberto Prado e Suzete Venturelli Fonte: Arquivo dos autores.

Mitômato (2023) é uma obra realizada pelo artista Fábio FON. O artista experimenta com o *ChatGPT*, entre os meses de fevereiro e março de 2023, perguntas sobre sua própria biografia. Nesse caso, a inteligência artificial *text-to-text* gera inúmeras informações incorretas. Compilado em um vídeo, as mentiras realizadas através das associações do robô demonstram o sentido do título atribuído a obra, reunindo os termos: “mitômano”, alguém que mente compulsivamente, com “autômato”, máquina que imita movimento de corpo animado (Figuras 10 e 11).



Figura 10 - Obra Mitômano (2023), de Fábio FON Fonte: Arquivo dos autores.



Figura 11 - Vista lateral da obra *Mitômano* (2023), de Fábio FON Fonte: Arquivo dos autores.

A série *Cartas do Inferno* (2023), criada pelo artista e pesquisador Chico Marinho, especialista em neurociência aplicada à computação criativa, cuja expertise experimenta uma variedade de linguagens artísticas, como mixmedias, pinturas, desenhos digitais e analógicos. Essa série tem como propósito dispor a criticar a relação profundamente intrincada entre a humanidade e avanço tecnológico. A utilização de diferentes materiais reflete essa conexão, enquanto as consequências desse entrelaçamento são representadas por seres demoníacos que transitam entre o mundo analógico e o digital (Figura 12).

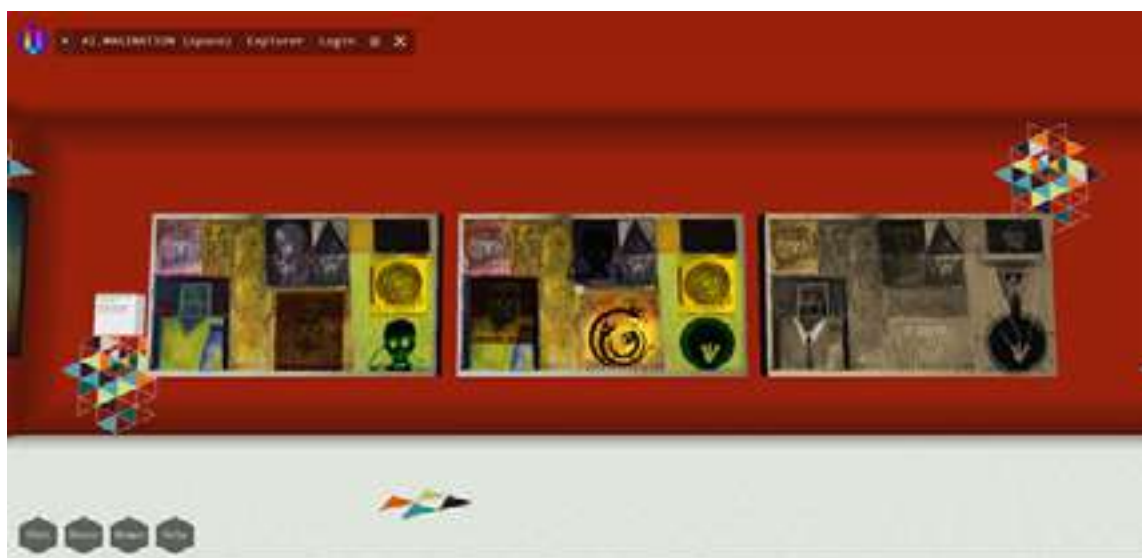


Figura 12 - Obra Série "*Cartas do Inferno*" (2023), de Chico Marinho Fonte: Arquivo dos autores.

A criação intitulada *Até agora, a violência, a guerra, a pilhagem, o assassinato e o roubo, etc. Foram aceitos como força motriz da história* (2021), concebida pelo artista Luiz Carlos de Oliveira Ferreira, são duas imagens generativas criadas através do *Google Colab* e projetos compartilhados na plataforma *Github*. Ambas as imagens revelam a constante presença da violência na sociedade e sua influência no progresso e na solidificação de uma cultura. A obra é o resultado da apropriação de imagens distorcidas que realça a violência e contextos violentos que compõem a narrativa histórica moldada pela humanidade, uma história que, muitas das vezes, negligenciam aqueles que perderam a vida em nome do progresso (Figura 13).



Figura 13 - Vista semi-lateral da obra *Até agora, a violência, a guerra, a pilhagem, o assassinato e o roubo, etc. foram aceitos como força motriz da história* (2021), de Luiz Carlos. Fonte: Arquivo dos autores.

Gustavo Machado, com sua série *Sem título* (2023), desdobra sua poética no campo do transumanismo. Neste trabalho, ele emprega o uso das ferramentas *Stable Diffusion* e *Wonder*, para criar três obras de arte. Estas obras exploram a concepção de memória coletiva, que emerge da hiperconectividade das redes neurais digitais presentes no cotidiano. Através do uso de inteligências artificiais, o artista retrata a expansão dessa memória compartilhada, destacando a habilidade generativa das ferramentas operadas. Isso evidencia a desmaterialização do corpo como um agente produtor de valor dentro do contexto do sistema capitalista (Figura 14).

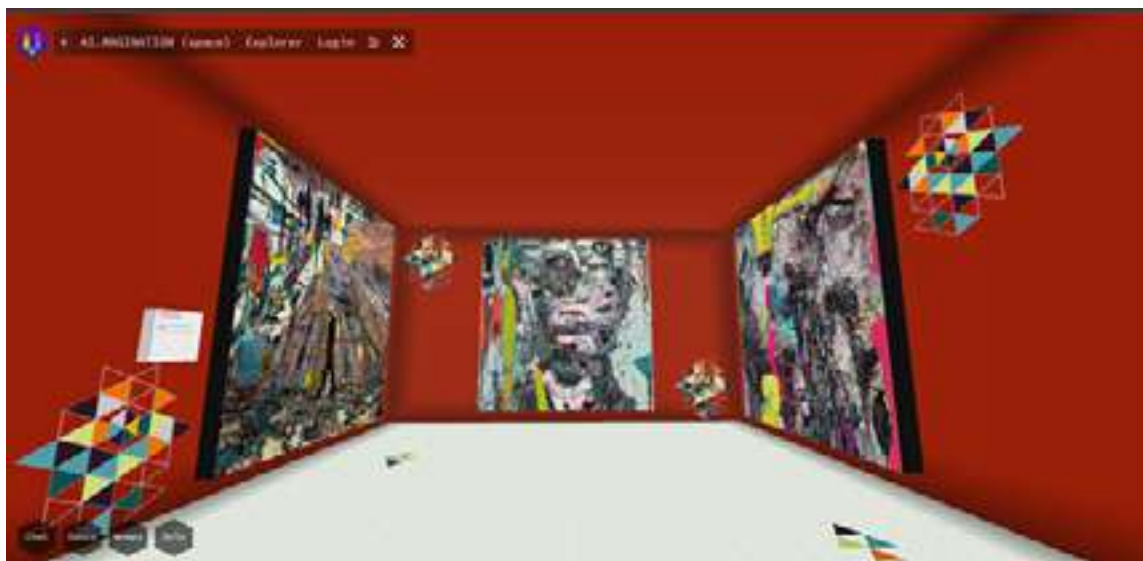


Figura 14 - Obra Sem título (2023), de Gustavo Machado Fonte: Arquivo dos autores.

A obra *Nova velha vida* (2023), concebida pelo artista Gabriel Rios, foi desenvolvida a partir do diálogo entre ele e o *ChatGPT*. Esse processo surgiu ao notar as análises críticas que o *chatbot* oferecia sobre a realidade social brasileira. A instalação artística tem como referência os “penetráveis” de Hélio Oiticica e incorpora diversos tipos de IA, como a conversão de texto em imagem, som e texto. A obra reúne quatro vídeos distintos e um poema em formato de áudio, escrito e recitado por um robô. O artista reconhece o potencial da máquina de criticar a realidade ao construir uma espécie de “puxadinho” para reforçar a situação da realidade brasileira, mesmo com o avanço tecnológico que é demonstrado pela própria IA (Figuras 15 e 16).



Figura 15 - Primeiro andar da obra Nova velha vida 1 (2023), de Gabriel Rios Fonte: Arquivo dos autores.

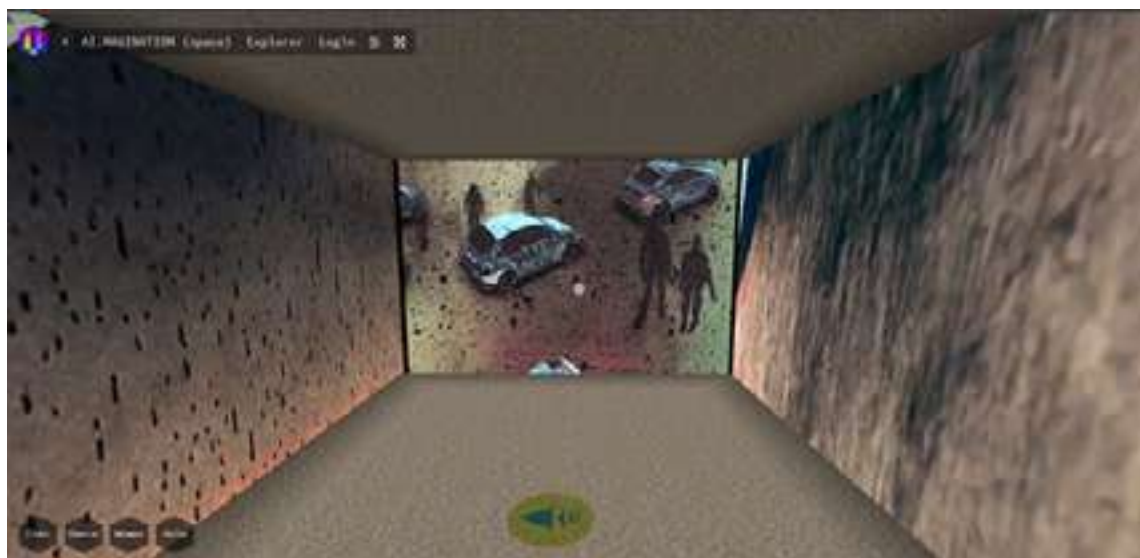


Figura 16 - Segundo andar da obra Nova velha vida 1 (2023), de Gabriel Rios Fonte: Arquivo dos autores.

Vinícius ‘Apos’ Viana, em sua criação denominada *Iconografias da Construção* (2023), localiza sua crítica no impacto das mineradoras em seus territórios estabelecidos para a exploração de minerais influenciado pelo avanço tecnológico. A crítica também reflete sobre a constante experiência de atualização e obsolescência que se baseia em uma lógica capitalista. A instalação apresenta ao público 4 videoartes geradas por inteligências artificiais *image-to-video*, treinadas com imagens aéreas de espaços convenientes para métodos de lavras e outras referentes à mineração. Os vídeos reforçam a intermitência da mineração, a construção e destruição do ecossistema (Figuras 17 e 18).



Figura 17 - Obra *Iconografias da Construção* (2023), de Vinícius “Apos” Viana Fonte: Arquivo dos autores.

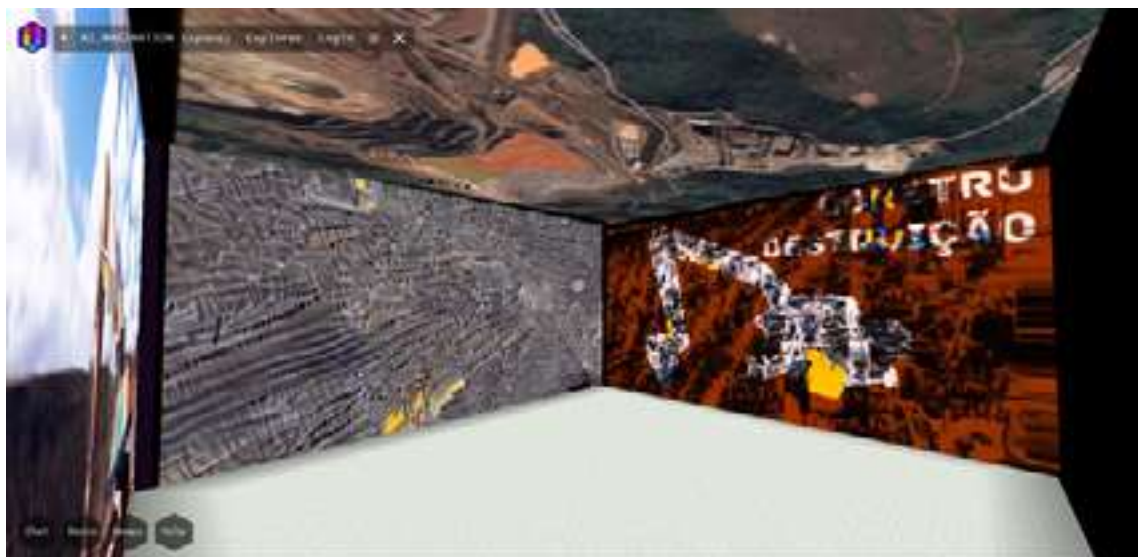


Figura 18 - Obra Iconografias da Construção e Destruição (2023), de Vinícius “Apos” Viana Fonte: Arquivo dos autores.

A obra intitulada *Brasília em um mundo pós-apocalíptico* (2023), criada pelo artista José Dias, constitui uma referência de acontecimentos anti-democráticos na capital federal brasileira ocorrido em janeiro de 2023. Neste trabalho, o artista opera o *Mid-journey* para gerar uma paródia das ações ocorridas na intenção de satirizar a lógica da economia política. Composta por 4 imagens que experimentam a mesma cena, o trabalho explora diferentes formas de devastação anti-democráticas (Figura 19).

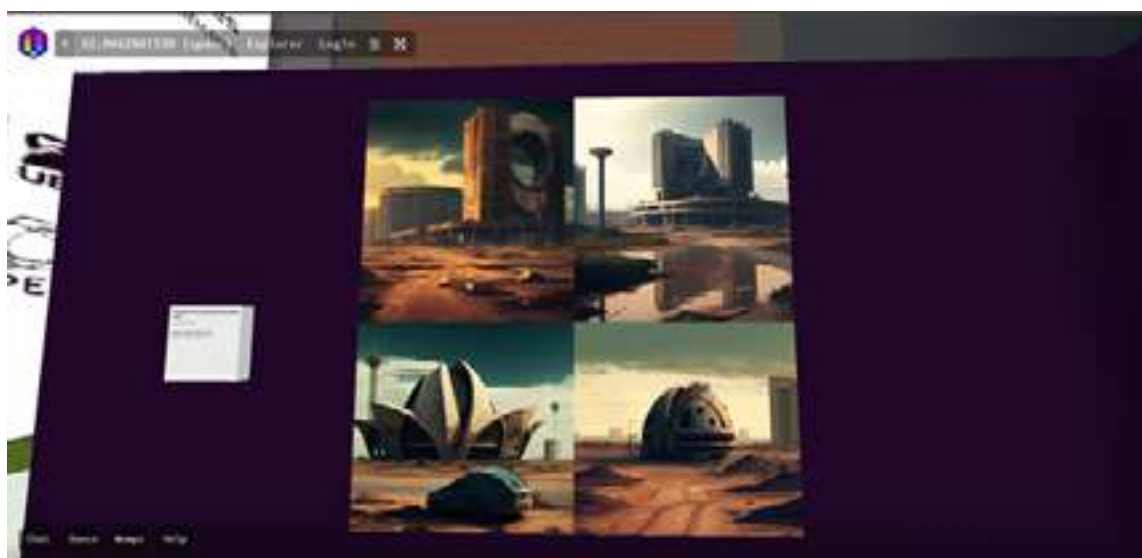


Figura 19 - Obra Brasília em um mundo pós-apocalíptico (2023), de José Dias Fonte: Arquivo dos autores.

A artista Gabriella Pedro utiliza a ferramenta *DALL.E 2* para gerar imagens através de uma descrição textual de fotografias do Aglomerado da Serra, região da cidade de Belo Horizonte/MG (Brasil) capturadas pela artista. A construção de uma realidade representada na obra *Rua lá perto da casa da sua avó* (2023) expressa o que foge das percepções desatentas. A obra provoca a reflexão e o reconhecimento de uma realidade que é interpretada pela IA (Figura 20).



Figura 20 - Obra Rua lá perto da casa da sua avó (2023), de Gabriella Pedro Fonte: Arquivo dos autores.

Na obra denominada *A.I. sem fio* (2023), o artista Eduardo Séllos relaciona a brincadeira infantil de telefone sem fio com o processo de tradução para diferentes linguagens através das inteligências artificiais. Exposta em tríptico, a criação provoca uma interlinguagem sinestésica. Ao gerar imagens a partir do *prompt* “Sinestesia”, o artista utiliza a ferramenta *image-to-audio*, que resulta em uma participação lúdica entre ferramentas inteligentes (Figura 21).



Figura 21 - Obra A.I. sem fio (2023), de Eduardo Séllos Fonte: Arquivo dos autores.

A obra *Interagindo com objeto não identificado* (2023), do artista Dera, reúne fotografias digitais de construções coloniais do interior de Minas Gerais. Exposta em formato de áudio, a criação convida para uma reflexão sobre a memória e a transformação pela qual somos sujeitos. O áudio é narrado com base em um texto gerado em uma conversa com o *ChatGPT*. Neste trabalho o autor convida o público para enviar imagens de objetos inanimados que compõem o seu cotidiano, propondo uma futura exposição com o material reunido (Figura 22).

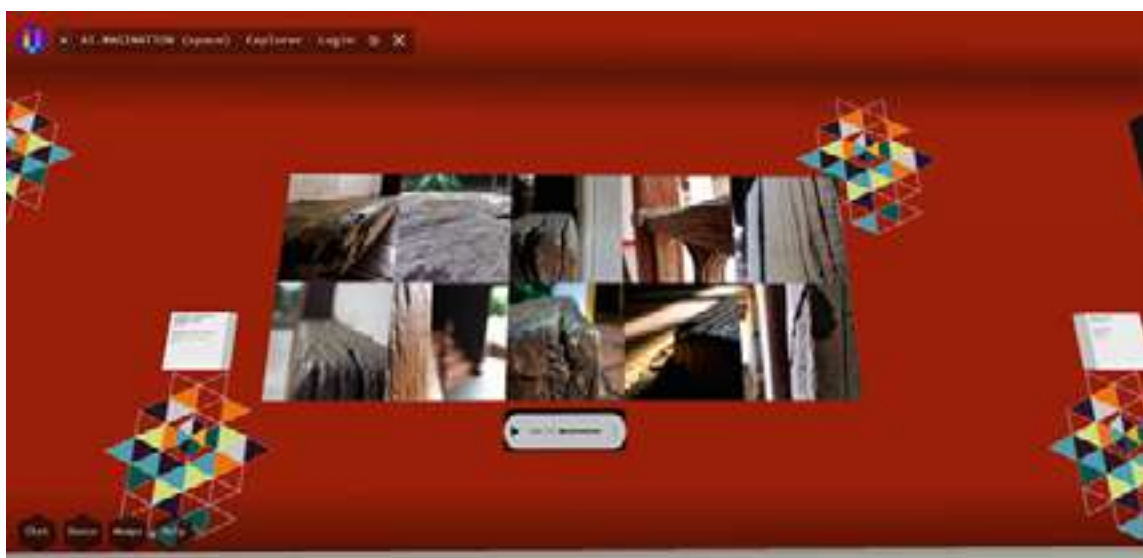


Figura 22 - Obra *Interagindo com Objeto mal identificado* (2023), de Dera Fonte: Arquivo dos autores.

A obra *Procura-se Helene Alberti* (2023), da artista Larissa Campello, trata sobre a cantora de ópera norte-americana que realizou experimentos ligados à aviação, mas que possui pouquíssimas informações referentes a sua biografia. Com o uso da ferramenta *text-to-image*, a criação forja encontros entre Helene Alberti e Santos Dumont. Paralelo a isso, a artista também tem um *website* dedicado à obra para que o público envie informações sobre Helene Alberti, como forma de construir a biografia da cantora, independente de sua veracidade (Figuras 23 e 24).



Figura 23 - Obra Procura-se Helene Alberti (2023), de Larissa Campello Fonte: Arquivo dos autores.



Figura 24 - Obra Procura-se Helene Alberti (2023), de Larissa Campello Fonte: Arquivo dos autores.

Skrif, em sua obra *Mosaico e releitura da obra Stanczyk de Jan Matejko* (2023), se apropria da pintura em formato digital e a transforma em um mosaico utilizando a ferramenta *AndreaMosaic*. Dessa forma, o trabalho apresentado é constituído por mosaicos da pintura que recria a pintura novamente. Metalinguística, a criação instiga a imersão do espectador, capaz de observar as etapas processuais da obra (Figuras 25 e 26).



Figura 25 - Obra Mosaico e releitura da obra Stanczyk de Jan Matejko (2023), de Skrif Fonte: Arquivo dos autores.



Figura 26 - Obra Mosaico e releitura da obra Stanczyk de Jan Matejko (2023), de Skrif Fonte: Arquivo dos autores.

A artista Ana Luiza Pedrosa Camilo, em sua criação intitulada *Consciência alada* (2023), explora a dimensão da consciência de animais ditos “irracionais”. Operando o *Midjourney*, a artista descreve para a ferramenta uma vivência que retrata a manifestação do raciocínio de um pombo ao se alimentar no contexto urbano. Dessa forma, a obra apresenta 2 imagens geradas a partir de *text-to-image*. A instalação é composta por uma escultura digital tridimensional em *voxel* (Figura 28) construída no *software* MagicaVoxel. Com essa composição, a obra reflete as condições e formas de vida de outras espécies que compartilham o planeta com os seres humanos. O uso de IA no processo de criação da obra revela parâmetros da imaginação algorítmica da ferramenta inteligente, que tenta representar e compreender a experiência do pombo (Figuras 27 e 28).



Figura 27 - Obra *Consciência Alada* (2023), de Ana Luiza Pedrosa Camilo Fonte: Arquivo dos autores.



Figura 28 - Obra *Consciência Alada* (2023), de Ana Luiza Pedrosa Camilo Fonte: Arquivo dos autores.

Considerações finais

Neste artigo tivemos como objetivo apresentar a exposição AI.MAGINATION. A nossa base teórica foram as referências vistas aqui e outras que compuseram a bibliografia das disciplinas de “Ateliê Transartes: hibridismos e apropriações artísticas em realidades diversas” e “Curadoria em arte digital”, mas que não foram abordadas devido à necessidade de objetividade neste texto.

O uso de IA na curadoria da exposição abriu caminho para estabelecer a dimensão criativa e conceitual do projeto, onde as capacidades humanas e da máquina se entrelaçam para construir expressões híbridas. Da perspectiva da produção artística, a combinação de diferentes linguagens artísticas, tecnologias e mídias permitiu aos discentes explorar novas possibilidades de expressão e experimentação poética, estendendo essa experiência até o público que visitou a exibição.

Ao longo deste trabalho vimos como o avanço tecnológico continua a moldar as percepções dos seres humanos e a maneira como o relacionamos com a arte. A expo-

sição AI.MAGINATION é um exemplo concreto de como as inteligências artificiais podem ser incorporadas ao campo artístico para enriquecer e expandir as fronteiras entre a arte, ciência e tecnologia. Compreendemos que, ao propor uma exposição que prevê o uso de IA em todas as frentes do projeto, pode ser desafiador lidar com essas ferramentas devido às suas complexidades e possibilidades. Sobretudo quando essa perspectiva é aplicada no programa das disciplinas ofertadas para o ensino superior em artes plásticas.

No que se refere ao processo criativo da AI.MAGINATION, trabalhar com ferramentas inteligentes simplificadas e abertas evidenciou, por parte dos artistas e membros da equipe curatorial, uma postura crítica a esses tipos de sistemas. Assim como as vanguardas do século XX apropriaram elementos da cultura e da própria história da arte para subvertê-los e construir novos caminhos, a exposição AI.MAGINATION incorpora e discute essas ideias na teoria e na prática poética. Ao utilizar inteligências artificiais para recontextualizar elementos do mundo digital, físico ou não, cria-se uma exposição que abriga obras de arte digitais que desafiam as expectativas do público e o *status quo* da arte em um contexto, muitas vezes, de críticas superficiais à IA.

Para concluir, a AI.MAGINATION convida a todos a explorar a união entre a imaginação e a inteligência, e a refletir sobre como essa junção pode nos levar a novas perspectivas sobre a realidade. As obras expostas são o resultado de um recorte criativo que considera a teoria e história da arte relembando nela as dimensões da apropriação, do hibridismo e da multiplicação de realidades.

Referências

BENJAMIN, Walter. *A obra de arte na era da sua reprodutibilidade técnica*. In: BENJAMIN, Walter. *Magia e Técnica, Arte e Política: Ensaios sobre literatura e história da cultura*. ROUANET, Sergio Paulo (Trad.). 3.ed. São Paulo: Brasiliense, 1987. p. 165-196.

BERGER, Peter L. LUCKMANN, Thomas. *A Sociedade como Realidade Objetiva*. In: _____. *A construção social da realidade: tratado de sociologia do conhecimento*.

FERNANDES, Floriano Souza (Trad.). 23^{ff} ed. Petrópolis: Editora Vozes, 1985. p. 69-126.

BULHÕES, Maria Amélia. *Lacunas como ponto de partida*. In: _____. *Artes plásticas no Rio Grande do Sul: pesquisas recentes*. Porto Alegre: Editora da UFRGS, 1995, v. 1, p. 113-128.

DEBORD, Guy; WOLMAN, Gil. *A user's guide to détournement*. In: _____. *Complete cinematic works: scripts, stills, documents*. Tradução de Ken Knabb. Oakland, CA: AK Press, 2003. p. 207-210.

MARMO, Alena Rizi.; LAMAS, Nadja de Carvalho. *O Curador e a Curadoria*. In: _____. *Investigações sobre arte, cultura, educação e memória*. Joinville: UNIVILLE, 2012. p. 19-29.

NUNES, Fábio Oliveira. *Como criar uma mentira de artista?*. In: GOBIRA, Pablo (Org.) *Percursos contemporâneos: realidade da arte, ciência e tecnologia*. Belo Horizonte: Editora da Universidade do Estado de Minas Gerais, 2018. p. 55-79.

RUPP, Betina. *Curadorias na arte contemporânea: considerações sobre precursores, conceitos críticos e campo da arte*. 240f. Dissertação (Mestrado em Artes Visuais) - Programa de Pós-graduação em Artes Visuais, Instituto de Artes da Universidade do Rio Grande do Sul. Porto Alegre, 2010.

RUPP, Betina. *O curador como autor de exposições*. *Revista-Valise*, Porto Alegre, v.1, n.1, p. 131-143, 2011.





■ Giselle Beiguelman, frames do vídeo generativo da série Flora Rebellis, versão 2024.
Imagens geradas com Inteligência Artificial (Style Gans)

■ [Acesse o vídeo clicando na imagem](#)





PARALELOS ENTRE AS INTELIGÊNCIAS NATURAL E ARTIFICIAL A PARTIR DE UM COMPARATIVO EPISTEMOLÓGICO:

limitações e impactos

DOI: <https://doi.org/10.35699/2965-6931.2023.47680>

LUIZ JOAQUIM NUNES*; MARIANA INÉS GARBARINO**

RESUMO: O estudo aborda a relação entre inteligência natural e artificial, destacando as limitações da IA atual em comparação com a humana. O trabalho realiza um comparativo epistemológico tomando como referencial a epistemologia genética piagetiana e o ethos humano. Analisam-se alguns aspectos da experiência humana que se acentuam como tais limitações: a afetividade, a moralidade, a criatividade e a ética. A afetividade humana desempenha um papel fundamental sobre as conclusões às quais a cognição humana chega. A IA carece da capacidade de desenvolver uma noção cognitiva de si mesma e, portanto, de moralidade, e a forma como é concebida não permite que novas estruturas cognitivas sejam produzidas para viabilizar uma criatividade completa. O entendimento sobre a ética humana contribui também para o entendimento dos impactos da IA. A análise, por fim, indica cenários para a aplicação de inteligência artificial, de acordo com suas limitações em cada caso.:

PALAVRAS-CHAVE: inteligência artificial; psicologia social; psicologia cognitiva; ética da tecnologia; futuro do trabalho.

Parallels between natural and artificial intelligence from an epistemological comparison: Limitations and impacts

ABSTRACT: The study addresses the relationship between natural and artificial intelligence, highlighting the limitations of current AI compared to human intelligence. The work involves an epistemological comparison grounded on Piagetian genetic epistemology and the human ethos. Some aspects of the human experience that are accentuated as such limitations are analyzed: affectivity, morality, creativity and ethics. Human affectivity plays a fundamental role in the conclusions reached by human cognition. AI lacks the capacity to develop a cognitive notion of itself and, therefore, of morality, and the way it is conceived does not allow new cognitive structures to be produced to enable complete creativity. Understanding human ethics also contributes to understanding the impacts of AI. The analysis, finally, indicates scenarios for the application of artificial intelligence, according to its limitations in each case.

KEYWORDS: artificial Intelligence; social psychology; cognitive psychology; technology ethics; future of work.

* Universidade de São Paulo / ** Universidade Federal do ABC (UFABC)

Introdução

A ideia de inteligência artificial (IA) não é um fenômeno recente, mas sim algo que permeia a história humana ao longo de várias culturas (Nunes, 2023). Desde antigas representações míticas de seres criados a partir do barro até algumas visões religiosas sobre a criação da vida, encontramos indícios de construções artificiais com atributos inteligentes. Essas narrativas, prévias à era dos computadores, sugerem a fascinação ancestral pelo conceito de inteligência semelhante à humana.

No entanto, à medida que a IA emerge como uma tecnologia proeminente na sociedade contemporânea, o próprio termo “inteligência artificial” pode ser aplicado de forma imprecisa, gerando ambiguidade em sua definição e levantando questões relevantes para campos como a comunicação científica e a psicologia social. Portanto, é válido explorar as raízes psicológicas que historicamente inspiraram o desenvolvimento da IA e refletir sobre a relação entre aprendizado de máquina e aprendizado humano. Esta investigação não apenas esclarece conceitos fundamentais, mas também destaca a importância contemporânea de delinear fronteiras conceituais nesse campo em constante evolução.

Nos seus estudos sobre cognição e inteligência, Jean Piaget (1960, 2005) explorou profundamente as facetas do desenvolvimento humano desde a infância, e sua obra permite uma compreensão mais clara sobre o quanto tais facetas são ou não identificáveis em inteligências artificiais também. Mais ainda, sua obra permite uma aproximação mais robusta da psicologia para a discussão de tais tecnologias que, embora não se tratem de inteligências numa perspectiva biológica, foram produzidas como a versão artificial daquilo que referenciam. Essa mesma referência pode ser revisitada e tomada como fundamento para discussão, após décadas de desenvolvimento tecnológico, sobre o quão distantes essa transversalidade as tornou.

Piaget define a inteligência como um estado de equilíbrio resultante das interações entre adaptações sensório-motoras e cognitivas, bem como acomodações e assimilações entre o organismo e o ambiente (Macedo, 1980). Ele descreve três formas de inteligência: ritmos, regulações e operações, que evoluem desde padrões básicos até processos complexos de raciocínio.

Essa estrutura de inteligência definida por Piaget encontra paralelos na definição da inteligência artificial apresentada por Stuart J. Russell e Peter Norvig. Eles propõem quatro abordagens para a IA: sistemas que agem como humanos, sistemas que pensam como humanos, sistemas que pensam racionalmente e sistemas que agem racionalmente (Russell & Norvig, 1995).

Um paralelo epistemológico pode ser estabelecido da seguinte forma: os ritmos correspondendo aos algoritmos principais da IA, as regulações correspondendo ao aprendizado de máquina baseado em feedback e otimização, e as operações correspondendo a aplicações específicas para resolver problemas, como inferência, ordenação e escrita.

Nesse paralelo as definições de inteligência de Piaget e as abordagens de IA propostas por Russell e Norvig apresentam uma congruência que permite mapear mais robustamente suas semelhanças e diferenças - a estrutura das três formas de inteligência se alinha de maneira consistente com as diferentes abordagens da IA. E então, tendo em vista onde essas semelhanças e diferenças possam ter maior relevância sobre a segurança de uma aplicação de IA, pode se fazer mais viável discutir impactos esperados dessas aplicações. Mais ainda, esse comparativo pode viabilizar que a psicologia se envolva com maior rigidez epistemológica em análises e debates sobre inteligência artificial, sua aplicação e regulação.

Sobre o papel da afetividade

Em seminários proferidos na Sorbonne (1953-54), Piaget enfatizou as conclusões de sua investigação sobre as relações de interdependência entre a cognição e a inteligência, mostrando o papel desempenhado pela afetividade em sua evolução e manifestação. A afetividade é a mola propulsora ou inibidora da ação, na forma de interesse ou evitamento (Piaget, 2005). De acordo com ele, a afetividade pode ser identificada em determinados cenários, com diversas consequências potenciais, podendo influenciar a execução de uma atividade tornando-a mais ágil, mais desafiadora ou mesmo mais intrincada.

A afetividade humana possibilita à pessoa reconhecer emoções em outros com base nas sensações que essas emoções evocam nela mesma, como é o caso da capacidade de identificar a dor que alguém possa estar experimentando (o que comumente é conhecido como “empatia” e que é atualmente estudado pela Teoria da Mente). Quando alguém se depara com algo detectado pelos sentidos, a compreensão não se limita apenas a um enfoque racional e descritivo; também abrange uma dimensão afetiva, ou seja, a forma como esse algo influencia emocionalmente a pessoa. Além disso, a afetividade não desempenha somente um papel transitório durante a interpretação desse algo. Mesmo que a racionalidade seja capaz de descrever e interpretar logicamente o que é percebido pelos sentidos, as sensações emocionais permanecem integradas à percepção global do indivíduo em relação àquele algo. Portanto, para Piaget (2005), a dialética afeto-cognição impregna cada ação, representação e produção humana.

Ao supostamente desenvolvermos uma inteligência artificial com o propósito de reconhecer emoções em indivíduos (Russell & Norvig, 1995), geralmente seguimos um processo específico: a máquina recebe acesso a um conjunto de imagens de pessoas, onde cada imagem já possui uma emoção atribuída. Ao receber uma nova imagem, a máquina busca avaliar o grau de semelhança entre essa nova imagem e as imagens existentes na base de dados, resultando na atribuição da mesma emoção associada às imagens consideradas mais parecidas. A definição de “semelhança” entre imagens pode ser estabelecida de diversas maneiras, levando em consideração desde as paletas de cores utilizadas na imagem até as estruturas geométricas identificadas nas características faciais das pessoas. Essas lógicas para mensurar a “semelhança” são pré-configuradas como parte do sistema de inteligência artificial. Portanto, é possível que uma máquina seja figurativamente capaz de discernir emoções em pessoas por meio de suas imagens, mas essa habilidade é estritamente resultado de cálculos matemáticos aplicados às imagens, sem envolver elementos próximos à afetividade humana ou sequer uma compreensão real sobre emoções e contextos afetivos.

Essas considerações sobre a dialética inteligência-afetividade podem, assim, indicar uma ampla gama de processos sutis e complexos que as inteligências artificiais, nos moldes atuais, não seriam capazes de executar. Essas distintas situações são de grande relevância para identificar os papéis que a IA poderia não ser capaz de desempenhar plenamente na sociedade humana.

Sobre a moralidade

De acordo com Piaget, ao longo da infância se desenvolve o paulatino processo de construção cognitiva social e afetiva do mundo, um mundo de objetos, normas e relações socio-afetivas com pessoas próximas do contexto familiar e das instituições que a criança frequenta (Piaget & Inhelder, 1974). Durante esse processo, as noções de objeto, espaço, tempo e causalidade são progressivamente desenvolvidas. Piaget observa que quando a criança demonstra a construção da noção de objeto, isso também sinaliza a formação da noção de si mesma: nos primeiros estágios da vida, a criança não revela - pelo menos não de maneira clara - uma compreensão de que os objetos continuam a existir independentemente de sua percepção individual: é o que na teoria piagetiana foi nomeado como noção de “objeto permanente”. No entanto, em algum ponto posterior, essa compreensão se torna evidente. É nesse ponto que a noção de si mesmo se torna perceptível do ponto de vista cognitivo, à medida que a criança se posiciona no universo como um objeto entre outros objetos, processo também conhecido como “descentração”, em detrimento do egocentrismo que tornava a criança o centro das referências para entender o mundo.

Essa compreensão também capacita o indivíduo a reconhecer outros indivíduos de maneira semelhante a si próprio, o que progressivamente possibilita a construção de sua própria identidade a partir de perspectivas culturais e morais diversas e que a transcendem. A pesquisa de Piaget sobre a moralidade, que foi posteriormente continuada por Kohlberg (Duska & Whelan, 1994), Turiel (Turiel, 1983) e, no Brasil, por Yves De La Taille (La Taille, 2002), inicialmente define a moral como um sistema de regras que emergem por meio da interação social. No entanto, uma leitura reducionista dessa abordagem desembocou em críticas por ser excessivamente racional. A partir desse ponto, surgem estudos mais focados no aspecto afetivo da moralidade, e o próprio Piaget (2005), especialmente ao longo de suas obras da maturidade, demonstra atribuir uma grande importância à dimensão afetiva na compreensão da moral.

Baseando-se no si mesmo cognitivo e na afetividade, um indivíduo teria a capacidade de alinhar suas necessidades, interesses e perspectivas com outros indivíduos,

resultando na construção de uma noção de si mesmo moral. O si mesmo moral é constituído conforme uma escala de valores e ideais construídos com base nas próprias experiências subjetivas e no exercício de uma moral autônoma permeada por contextos socioculturais que transmitem e perpetuam determinados valores morais. Em outras palavras, tanto a compreensão cognitiva de si mesmo e do contexto social quanto a afetividade (o “querer ser moral”) desempenham um papel crucial na formação do si mesmo moral, e, conseqüentemente, na constituição da moral em si.

Conforme discutido na seção sobre a afetividade, as máquinas em sua forma atual são incapazes de processar afetividade humana, uma vez que isso está fora do âmbito de suas naturezas essencialmente racionais. Além disso, as máquinas apresentam uma distinção cognitiva notável em relação aos seres vivos: o software e o hardware, duas partes que operam de maneira conjunta, mas não são exclusivamente interdependentes. Se um software é transferido para outro hardware completamente idêntico, a inteligência artificial não sofre alterações, já que todos os seus componentes, recursos e capacidades permanecem, em essência, inalterados. Da mesma forma, se o software ligado a um determinado hardware é substituído por outro software idêntico, a mesma lógica é aplicável. Essa característica torna desafiador determinar claramente onde se inicia e onde se encerra o conceito de “si mesmo” dentro da inteligência artificial, pelo menos quando tentamos aplicar a mesma definição que usamos para os seres humanos.

Mesmo que uma IA possua um software capaz de reconhecer fisicamente sua localização e a composição do hardware a que está conectada, ela não possui fundamentos lógicos nem capacidade cognitiva para sentir medo em relação à sua própria destruição, a menos que seja explicitamente construída e programada para fazê-lo, como ilustrado pelas Leis da Robótica propostas por Asimov (Asimov, 1941).

Em outras palavras, uma máquina, por si só, carece da capacidade de desenvolver uma percepção cognitiva de si mesma e, portanto, de construir uma moralidade. Uma abordagem alternativa seria a implementação de um conjunto de regras que governaria as conclusões e ações da máquina, de maneira similar às Leis da Robótica e à concepção inicial de moralidade delineada por Piaget, que excluiria a dialética afetividade-inteligência, na tomada de decisões morais. Embora esse sistema possa funcionar de maneira distinta em relação ao modo humano, ele poderia desempenhar um papel de

condição de contorno (Chvátal, 1983) no contexto das interações entre seres humanos e máquinas.

Existe ainda um problema amplamente abordado quando se discute a tomada de decisões em cenários envolvendo dilemas morais, possivelmente o problema mais conhecido nessa categoria: o dilema do bonde. Este dilema, introduzido por Foot (1967) e aprofundado por Thompson (Thompson, 1976), lida com situações nas quais se aplica o princípio conhecido como “duplo efeito”. Em resumo, surge a necessidade de tomar uma decisão dentro de um conjunto de opções disponíveis, em que cada alternativa carrega simultaneamente consequências percebidas como positivas e outras, como negativas. No caso do dilema do bonde, a situação envolve um bonde em movimento que deve optar por uma de duas vias possíveis, cada uma delas contendo um grupo distinto de pessoas, as quais seriam fatalmente atingidas.

Além de representar um dilema quanto à escolha da decisão a ser tomada, o dilema também aborda a situação em que a decisão é feita e quanto ela pode influenciar a percepção humana das consequências e, conseqüentemente, a própria tomada de decisão. Portanto, a tomada de decisões por parte das pessoas pode revelar inconsistências conforme avaliadas pelas suas consequências.

Ponderar sobre esse princípio poderia se mostrar relevante ao considerar a possibilidade de delegar a uma inteligência artificial a tomada de decisões de natureza moral. Se, entre seres humanos, pode não haver um consenso claro quanto à melhor opção a ser escolhida, qual seria o impacto na convivência entre humanos se máquinas fossem responsáveis por tais decisões? Além disso, devido à sua limitada compreensão moral, a máquina poderia encontrar dificuldades em reconhecer a presença de consequências negativas e positivas que devem ser ponderadas. Talvez, como debatido na seção anterior, a máquina pudesse até discernir algumas implicações negativas - como, em um cenário hipotético, a possibilidade de uma pessoa se ferir devido a uma ação da máquina. No entanto, esse discernimento não seria fundamentado em um julgamento moral, mas sim em uma avaliação lógica ou algébrica, a qual precisaria ser previamente programada na máquina.

Sobre a criatividade

Quando exploramos a efetiva implementação da inteligência artificial na realidade, expondo suas limitações e os devidos cuidados, em vez de meramente enfatizar as fascinantes possibilidades futuras da tecnologia, a obra “What Computers Can’t Do: The Limits of Artificial Intelligence” (Dreyfus, 1972) se destaca como uma das mais renomadas, e Dreyfus emerge como um crítico proeminente da inteligência artificial ao longo do último século (Dreyfus, 1972; 1991; 2007).

Tal como mostrado por Gomes (2018), Dreyfus expõe sua experiência junto ao departamento de engenharia do MIT durante o estudo sobre inteligência artificial realizado pelo mesmo. Utilizando as bases estabelecidas por Heidegger (Heidegger, 1927), destaca que a abordagem empregada pelo departamento apresentava certos aspectos suscetíveis a questionamento. A questão central de sua crítica reside no fato de que a inteligência artificial, desde suas origens, opera com base em um conjunto de representações e operações previamente definidas, limitando-se às estruturas que estão incorporadas em sua ontologia. Por exemplo, uma IA que produz textos, como o Chat-GPT, pode até produzir diferentes tipos de textos, mas opera apenas dentro do âmbito dos textos, processando palavras, frases ou sílabas, sem em geral uma capacidade de processar imagens ou matrizes - e sem a necessidade de tal capacidade realmente.

Isso resulta em uma separação entre a máquina e o ambiente em que ela opera, já que sua habilidade de processar qualquer informação que recebe está condicionada a essas estruturas preestabelecidas. Esse cenário difere do que ocorre com os seres humanos: antes de compreendermos o mundo ao nosso redor, compreendemos a nós mesmos como parte intrínseca desse mundo, de maneira integrada e, inicialmente, sem distinções claras. Nos primeiros estágios do desenvolvimento humano, o eu e o entorno são confundidos, e por momentos fusionados, criando uma indistinção entre o que é produção do sujeito (o que do si mesmo) e o que é emergente do mundo dos objetos ou das pessoas próximas. Progressivamente, ao longo do desenvolvimento, passamos a identificar, com maiores precisões, as distinções entre os outros e o si mesmo, no âmbito das crenças, dos valores, dos pensamentos, das ações e da tomada de consciência sobre as próprias decisões (Piaget & Inhelder, 1974).

A base pré-existente ao desenvolvimento da cognição artificial em uma máquina

permanece como um pressuposto para essa cognição, resultando em sua construção não a partir do mundo, mas em direção a ele e exclusivamente a ele - tal cognição não seria sustentável em um contexto diferente. A máquina, concebida como uma ferramenta para os seres humanos, não opera sobre o mundo a partir de sua própria existência nele, mas sim a partir da existência dos seres humanos que a criaram. A absorção de novas informações do mundo pode influenciar uma inteligência artificial a modificar a maneira como realiza uma tarefa, porém, não altera as capacidades específicas que ela é capaz de executar. Existem paradigmas por hora inerentes ao funcionamento da máquina, quando levamos em consideração a natureza atual da tecnologia e nossa capacidade de avaliá-la. Esse aspecto da inteligência artificial, presente em projetos de IA como um todo, é, resumidamente, o que motiva Dreyfus a criticamente se referir a ela como *inteligência artificial heideggeriana*.

De maneira potencialmente expansível ao infinito, a incorporação progressiva de estruturas na máquina pode facultar operações algébricas de crescente complexidade. É justamente esse fenômeno que Dreyfus aborda em sua tese, na qual sugere que a solução para a chamada IA heideggeriana reside no contínuo aprimoramento desse caráter heideggeriano.

Desse modo, o tipo de inteligência artificial que conhecemos tem sua origem na existência prévia de uma cognição, ao contrário de construir essa cognição a partir de sua interação com o mundo, como ocorre com os seres humanos. Além disso, a existência pré-estabelecida dessa cognição por meio de sua ontologia interna pode permitir algum grau de percepção do mundo, embora não possibilite a expansão das próprias capacidades cognitivas. Consequentemente, a compreensão significativa do mundo e dos outros permanece limitada, mesmo com qualquer processo de aprendizado, visto que, na prática, a máquina apenas é capaz de compreender algo em um nível representativo, definindo algo com base nas informações disponíveis, em vez de uma experiência própria referente àquilo.

No contexto deste estudo, e diante do que foi até aqui exposto, propomos então entrelaçar a crítica de Dreyfus com a psicologia de Piaget. Ainda que as estruturas subjacentes à inteligência natural e à inteligência artificial possuam similaridades estruturais, Piaget (1960), e posteriormente os trabalhos de Macedo (1980) exploram essas estruturas como estágios do desenvolvimento. Quando se pensa no desenvolvi-

mento da inteligência segundo Piaget, passando pelos ritmos, regulações e operações, é possível estabelecer uma analogia entre essa sequência e a trajetória da inteligência artificial, a partir dos circuitos mais gerais, passando pela implementação de códigos e sua generalização. Contudo, é importante destacar que essa analogia não se aplica às entidades concretas de inteligência artificial, como máquinas ou softwares específicos. Ou seja, a tecnologia não apresenta por conta própria ritmos, regulações e operações ao longo do tempo como ocorre no contexto da psicologia genética piagetiana. A analogia entre a inteligência piagetiana e a inteligência artificial é viável apenas porque a própria IA foi concebida desde o início como uma analogia, e mesmo na atualidade essa analogia continua válida.

Ao levar isso em consideração, é possível notar que as estruturas propostas por Piaget nos auxiliam a discernir entre diferentes categorias de máquinas e inteligências artificiais. Contudo, a conceituação do desenvolvimento infantil, particularmente em relação à etapa sensório-motora e à função simbólica, não somente não se aplica às máquinas, como também se apresenta como uma sequência inversa à construção de uma inteligência artificial. Nesse contexto, a cognição inicial é implementada antes de qualquer interação com o mundo físico, constituindo uma ordem oposta àquela presente no desenvolvimento infantil.

Essa discrepância assume importância ao considerarmos a viabilidade das máquinas resolverem problemas que ainda permanecem insolúveis para os seres humanos. Devido à sua operação, toda inteligência artificial demanda algum grau de instrução originado por um ser humano, e as estruturas artificiais que adquirem inevitavelmente delimitam suas capacidades. Em outras palavras, as estruturas cognitivas possíveis em uma inteligência artificial formam necessariamente um subconjunto das que os seres humanos já foram capazes de conceber, estabelecendo assim um limite para a artificialização que continuará a perdurar no futuro, pelo menos enquanto os computadores mantiverem seu funcionamento atual.

Portanto, qualquer definição de criatividade que transcenda a simples recombinação de elementos conhecidos encontra-se fora do alcance de uma IA. Essa recombinação, dependendo do contexto, pode ser encarada como uma forma simples de criatividade, inclusive observada no desenvolvimento de crianças também. A produção

de novas estruturas e conceitos, porém, possibilita uma criatividade mais generalizada em seres humanos, ainda inviável na implementação de inteligências artificiais.

Sobre a ética

Após explorarmos os tópicos da afetividade, da moral e da criatividade, deparamo-nos com outro conjunto de assuntos altamente significativos para compreender as limitações e as repercussões da IA: a ética. Um termo que adquire distintas definições, variando conforme o contexto em que é empregado e a área em que é estudado, mas que, provavelmente, em todas as suas instâncias guarda alguma ligação com o objeto deste estudo.

Ribeiro propõe que a essência subjacente a tudo que é englobado pela palavra “ética”, a base de qualquer situação que possa ser identificada como um “dilema ético”, corresponderia aos significados que, de maneira etimológica e filológica, foram associados ao termo “ethos” (Ribeiro, 2018; 2019). Nesse enfoque, a autora estabelece uma definição para o *ethos humano* como “*uma relação natural (como que) de moldagem contínua e recíproca entre um lugar e a vida de seus habitantes*”. É importante esclarecer que, conforme a concepção da autora, a vida se manifesta de forma empírica tanto em âmbitos *comportamentais* quanto *subjetivos*.

Para uma apreensão mais profunda dessa definição, é possível iniciar a análise a partir de seus componentes fundamentais, conforme Ribeiro (2018):

- O *comportamento* diz respeito às ações humanas que estão sujeitas a contingências que podem resultar em sua supressão ou reforço, como é explorado, por exemplo, no campo da análise do comportamento;
- A *subjetividade* abrange todas as experiências que se desenrolam abaixo da superfície da pele, perceptíveis pelos indivíduos humanos, e é objeto de estudo em áreas como a psicanálise e a psicobiologia;
- O *lugar* abrange tanto o espaço físico quanto o relacional e simbólico, no qual ocorrem o comportamento e a subjetividade, sendo esse conceito objeto de estudo em disciplinas como a geografia e a arquitetura;
- No âmbito deste contexto, a concepção de *moldagem* diz respeito à interação recíproca entre os três elementos, resultando em modificações tanto em termos ma-

teriais quanto simbólicos. Essas alterações ocorrem de maneira conjunta ao longo do tempo, assemelhando-se, em sua estrutura, a uma série temporal na matemática, a um processo estocástico na estatística ou a um sistema de fluidos na mecânica. Uma modificação, mesmo que sutil, em qualquer um dos elementos poderia, como decorrência das relações éticas, exercer influências diversas sobre todos os três, incluindo uma retroação sobre o próprio elemento, à medida que o tempo avança.

Essa definição modela epistemologicamente o *ethos* como uma relação fundamental na experiência humana, sempre manifestando-se por meio de seus componentes, e também ressalta a natureza interdisciplinar do seu estudo. Isso implica que uma área de pesquisa focada exclusivamente em um dos elementos provavelmente oferecerá uma visão limitada da relação como um todo. Em paralelo, refletir sobre os impactos de uma aplicação de inteligência artificial apenas dentro de um dos elementos chega também a essa limitação.

Nessa perspectiva, quando um indivíduo executa um novo comportamento no contexto de um grupo ao qual pertence e esse comportamento é bem recebido, ocorre um reforço positivo. Esse reforço se manifesta tanto subjetivamente para o agente quanto por meio do aprendizado observacional por parte das outras pessoas no entorno, o que pode incrementar a probabilidade de sua repetição. Isso vale tanto para o próprio agente quanto para qualquer membro do grupo. Mesmo aqueles que inicialmente poderiam ter tido reservas ou desaprovação em relação ao comportamento podem gradualmente modificar sua avaliação, resultando em um impacto tangível sobre a subjetividade da pessoa em questão. Além disso, se esse comportamento influenciar o ambiente onde essas pessoas estão - como o ato de jogar lixo em uma rua, por exemplo -, o próprio lugar sofre uma alteração direta. De múltiplas maneiras, a relação ética - o *ethos* - está em constante movimento, ocorrendo e atualizando-se (Ribeiro, 2018), conforme exemplificado aqui.

Ao inserir uma inteligência artificial em um contexto humano, a tecnologia tem o potencial de influenciar tanto o comportamento humano quanto a subjetividade e o lugar, de acordo com suas características e funcionamento específicos. Através do *ethos*, da conexão ética subjacente, existe sempre a possibilidade de afetar todos esses três elementos, bastando impactar pelo menos um deles. Portanto, torna-se imperativo considerar a inteligência artificial de uma perspectiva ética, ou seja, examinar as ques-

tões éticas que emergem da incorporação da IA em situações da vida real.

Nesse sentido, apresentaremos a seguir um exemplo prático que permite vislumbrar essa inter-relação entre os elementos que conformam essa complexa rede do ethos. Atualmente diversas ferramentas de inteligência artificial estão disponíveis para auxiliar empresas brasileiras na seleção de candidatos. Essas ferramentas variam em complexidade, desde as versões simples que aplicam filtros predefinidos, até as mais avançadas que buscam escolher candidatos com maior probabilidade de permanência, por exemplo. No entanto, existem riscos mais sutis nessas aplicações: ao priorizar candidatos com maior tempo de permanência numa empresa onde homens ficam em média mais tempo que mulheres, por exemplo, a lógica de redução de custos pode negligenciar fatores como a discriminação de gênero que pode estar por trás dessa diferença na permanência, levando a recomendações enviesadas. Em ambas as situações, a compreensão limitada da tecnologia pode levar a uma transferência indevida de responsabilidade para a IA, ignorando suas limitações e complexidades subjacentes.

Nesse exemplo é possível antecipar repercussões nos três elementos do ethos em questão:

- Caso a inteligência artificial tenda a selecionar homens como candidatos mais apropriados devido a um histórico de permanência mais longa na empresa, isso incentiva um desbalanceamento de gênero nas contratações. Mesmo que anteriormente a seleção fosse equilibrada em termos de números entre homens e mulheres, a proporção de candidatos escolhidos passa a ser predominantemente masculina. Isso tem um impacto direto sobre os comportamentos dos indivíduos encarregados do processo de recrutamento, possivelmente sem que eles percebam plenamente. No entanto, a questão se torna mais séria do que meramente influenciar os comportamentos, pois pode sugerir que a maior permanência de homens na empresa seja resultado de fatores discriminatórios, e a administração da empresa pode não ter consciência desse aspecto ao fazer uso da tecnologia. Isso significa que, ao focar somente no componente comportamental, o possível impacto social já assume uma natureza prejudicial e preocupante;
- Quando a tecnologia é empregada com o propósito de “selecionar os candidatos ideais para a empresa” - ainda que o critério inicial seja baseado na estimativa de tempo de permanência na organização -, emerge um descompasso entre o resultado gerado pela tecnologia e a interpretação desse resultado. Esse descompasso, no en-

tanto, pode fomentar a noção equivocada de que homens são intrinsecamente mais adequados para a posição, embora a definição de “candidato ideal para a empresa” não seja uniformemente compartilhada entre todos os envolvidos. Tal situação pode induzir percepções enviesadas de gênero entre os membros da empresa, impactando sua subjetividade por meio de suas concepções, opiniões e representações;

- Finalmente, existem dois ângulos pelos quais podemos examinar o impacto sobre o lugar. Em uma perspectiva mais restrita, focando exclusivamente na empresa em questão, ocorre uma transformação na composição da força de trabalho, com a proporção de homens tendendo a aumentar gradualmente ao longo do tempo por conta das sugestões da IA. Por outro lado, em uma perspectiva mais ampla, quando consideramos a colocação dessa empresa dentro de sua sociedade, a preferência pela contratação de candidatos homens influencia, ainda que inicialmente em pequena escala, as oportunidades de emprego para mulheres e homens, o que, por sua vez, intensifica as disparidades socioeconômicas entre os gêneros.

Portanto, uma tecnologia introduzida inadvertidamente em uma empresa que já apresentasse uma cultura organizacional discriminatória poderia, na verdade, acentuar ainda mais sua condição, abrangendo desde impactos individuais até implicações de alcance mais amplo.

A apresentação detalhada deste exemplo tem como objetivo destacar que a introdução da inteligência artificial em qualquer contexto pode, de acordo com a concepção do ethos conforme definida por Ribeiro, acarretar em efeitos éticos significativos (Ribeiro, 2018; 2019). Simultaneamente, esta estrutura de observação e interpretação do ethos pode servir como uma abordagem para antecipar de maneira mais precisa tais ramificações, facilitando a tomada de decisões mais cautelosas em relação à implementação da inteligência artificial. Isso é válido independentemente do cenário, com a ênfase na consideração do princípio da beneficência.

Três cenários possíveis para a ia

Ao considerarmos as restrições e repercussões da inteligência artificial abordadas até o momento, torna-se evidente a existência de diversos cenários para sua implementação. Alguns desses contextos envolvem a tecnologia executando tarefas de maneira análoga aos seres humanos, enquanto em outros casos isso não é aplicável, especialmente em situações que envolvem elementos afetivos, morais ou criativos.

Pode ser proveitoso, tanto para facilitar a análise de aplicações de inteligência artificial quanto para embasar futuras discussões, sugerir a adoção de uma terminologia específica para categorizar esses distintos cenários. Considerando, portanto, os tópicos abordados de forma integrada ao longo deste estudo, identificamos inicialmente três desses cenários:

- **Emulação:** consiste na reprodução artificial de uma operação por meio de passos análogos àqueles que um ser humano seguiria para executá-la, sem evidente perda cognitiva. Essa abordagem tende a ser viável em tarefas estritamente dependentes de raciocínio lógico e matemático, sem a necessidade de incorporar elementos como afetividade, moralidade ou criatividade para atingir seu objetivo. No entanto, é relevante considerar que, como mencionado anteriormente conforme a teoria piagetiana (Piaget, 2005), a afetividade sempre está entrelaçada à inteligência humana natural, e esse aspecto não seria reproduzido artificialmente. Exemplos podem abranger desde problemas matemáticos simples, como equações de primeiro grau, até tarefas mais complexas, como a automação da contabilidade empresarial. Nesse contexto, é possível argumentar que as máquinas têm potencial para executar essas tarefas de maneira mais rápida e com menor probabilidade de erro do que os seres humanos, devido à natureza lógica e matemática da máquina, que atende aos requisitos necessários para a realização, sem ser influenciada pela afetividade, que, nesses casos, não é uma consideração prioritária;

- **Simulação:** refere-se à realização artificial de uma operação por

meio de etapas diferentes daquelas que um ser humano seguiria, potencialmente resultando em um produto final suficiente, porém qualitativamente inferior em algum aspecto. Um exemplo é o reconhecimento de emoções em imagens de pessoas, conforme previamente mencionado. Nesse caso, uma IA pode até obter um resultado quantitativamente satisfatório para fins de automação, mas usando um método suscetível a falhas - e sem necessariamente um embasamento científico -, que um profissional humano capacitado e disponível não empregaria. Outra ilustração é a seleção de candidatos por meio de uma métrica que, em teoria, representa a “pessoa mais adequada para trabalhar na empresa”. Essa métrica é um modelo numérico consistente e embasado financeiramente, com foco na redução de custos. Entretanto, ela desconsidera questões cruciais da psicologia organizacional, sujeitando o negócio a se alienar de sua própria cultura. Em situações onde a afetividade, a moralidade ou a criatividade estão intrinsecamente ligadas ao problema a ser resolvido, a possibilidade de emulação provavelmente é limitada. No entanto, alternativas de simulação podem ser exploradas. Porém, a criação e implementação dessas alternativas requerem maior cautela devido à sua natureza artificial, que pode potencialmente causar impactos por meio do ethos, afetando a relação entre comportamento, subjetividade e o lugar em que ocorre.

- Abstenção: refere-se à impossibilidade de executar uma operação de maneira artificial, devido a limitações tecnológicas ou circunstanciais. Esses são os casos em que não existem alternativas de emulação ou simulação que produzam resultados adequados às necessidades e às condições específicas. Como resultado, a decisão tomada é de não aplicar inteligência artificial àquela situação.

É importante notar que há situações em que várias abordagens do problema são possíveis, e essas abordagens podem se encaixar em diferentes cenários entre os três mencionados. Isso ocorre porque os cenários se referem aos cursos de ação escolhidos, não apenas às situações específicas nas quais estão inseridos e nas quais se baseiam. A mesma operação pode ter possibilidades de emulação e simulação, e a escolha sobre

qual alternativa implementar cabe à decisão humana.

Ao considerar os impactos éticos de cada um dos três cenários, podemos tomar decisões mais informadas e socialmente responsáveis, incluindo a decisão de evitar a aplicação de inteligência artificial sempre que se constatar que a operação inteligente natural dos seres humanos na resolução de um problema envolve aspectos comportamentais e/ou subjetivos – especialmente afetividade, moralidade e criatividade – que não podem ser emulados ou simulados devido a limitações tecnológicas, sejam elas temporárias ou permanentes.

Considerações finais

Em um mundo onde a inteligência artificial tem se expandido e transformado diversas áreas da sociedade, é essencial reconhecer que há aspectos intrínsecos à inteligência natural humana que não podem ser completamente artificializados. A afetividade, a moralidade e a criatividade são elementos fundamentais da nossa cognição que desempenham papéis cruciais em nossa compreensão do mundo e na tomada de decisões. Embora a inteligência artificial possa emular ou simular algumas operações racionais, ela carece das dimensões afetivas, morais e criativas que são inerentes à nossa natureza e que nos permitem tomar decisões em contextos culturais específicos e complexos, permeados por elementos simultaneamente cognitivos e socioafetivos.

A afetividade humana permite-nos compreender emoções em outras pessoas e perceber o mundo de maneira subjetiva e emocionalmente dimensionada. A moralidade, construída sobre a base do si mesmo cognitivo e afetivo, guia nossas interações e decisões com base em princípios éticos. Além disso, a criatividade humana transcende regras e estruturas predefinidas, permitindo-nos explorar novas soluções e ideias.

A inserção da inteligência artificial em contextos humanos exige uma avaliação cuidadosa dos possíveis impactos e limitações. A tecnologia pode emular tarefas lógicas, mas sua incapacidade de compreender e processar emoções, discernir nuances éticas e criar de forma genuína requer uma abordagem responsável e cautelosa. Tomar decisões éticas sobre o uso da inteligência artificial é fundamental para evitar conse-

quências indesejadas e promover um progresso tecnológico alinhado com os valores humanos.

Em um cenário em que a IA possa ser emulada, simulada ou onde sua inserção é inviável devido a limitações, é crucial que a sociedade considere cuidadosamente os riscos e benefícios associados a cada abordagem. Ao delinear e compreender os três cenários aqui propostos, os tomadores de decisão podem optar por aplicar a IA de maneira mais responsável e atenta à ética, preservando a integridade das dimensões afetivas, morais e criativas que nos tornam humanos. Isso ajudará a assegurar que a inteligência artificial seja uma ferramenta que potencializa nossa capacidade de melhorar o mundo, ao invés de substituir completamente o que é, por essência, humano.

Referências

ASIMOV, Isaac. *Liar!*, em *Astounding Science Fiction*, edição de maio de 1941. Disponível em https://archive.org/stream/Astounding_v27n03_1941-05#page/n41/mode/2up. Acesso em: 7 set. de 2023.

CHVÁTAL, Václav. *Linear Programming*. New York: W. H. Freeman, 1983.

DREYFUS, Hubert. *Being-in-the-World: A Commentary on Heidegger's Being and Time, Division I*. Cambridge: MIT Press, 1991.

DREYFUS, Hubert. *What Computers Can't Do: The Limits of Artificial Intelligence*. 1972. Disponível em: <https://archive.org/details/whatcomputerscan017504mbp/page/n7/mode/2up>. Acesso em: 7 set. de 2023.

DREYFUS, Hubert. *Why Heideggerian Artificial Intelligence failed and how fixing it would require making it more Heideggerian*, em *Philosophical Psychology*. Londres, 2007.

DUSKA, Ronald & WHELAN, Mariellen. *O desenvolvimento moral na idade evolutiva: um guia a Piaget e Kohlberg*. São Paulo: Loyola, 1994.

FOOT, Philippa. *The Problem of Abortion and the Doctrine of the Double Effect*, na revista *Oxford Review*, número 5 Oxford: Basil Blackwell. 1967.

GOMES, Rodrigo Benevides Barbosa. *Hubert Dreyfus e Martin Heidegger: Representação e Cognição*, na revista *Kinesis*. Disponível em <https://revistas.marilia.unesp.br/index.php/kinesis/article/view/8070>. Acesso em: 7 set. de 2023.

HEIDEGGER, Martin. *Ser e Tempo*. Rio de Janeiro: Editora Vozes, 2005. Publicação original em 1927.

LA TAILLE, Yves de. *O Sentimento de Vergonha e suas Relações com a Moralidade*, em *Psicologia: Reflexão e Crítica*, edição 15, 13-25. Porto Alegre: Springer, 2002.

MACEDO, Lino de. *As estruturas da inteligência, segundo Piaget: ritmos, regulações e operações*. Rio de Janeiro: 1980.

PIAGET, Jean. *Inteligencia y afectividad*. Buenos Aires: Aique, 2005.

PIAGET, Jean. *Psychology of Intelligence*. New Jersey: Littlefield, Adams & Co., 1960.

PIAGET, Jean, & INHELDER, Bärbel. *A Psicologia da criança*. São Paulo: Difel, 1974.

RIBEIRO, Sandra Maria Patrício. *Lições preliminares para o estudo do ethos humano contemporâneo*. Tese de livre docência. São Paulo: 2018. Disponível em: <https://teses.usp.br/teses/disponiveis/livredocencia/47/tde-05042019-100757/pt-br.php> . Acesso em: 7 set. de 2023.

RIBEIRO, Sandra Maria Patrício. *Uma perspectiva para a compreensão do ethos humano*, em *Ethos Humano e mundo contemporâneo*. São Paulo: Baracoa, 2019.

THOMPSON, Judith. *Killing, letting die, and the trolley problem*, na revista *Monist: An International Quarterly Journal of General Philosophical Inquiry*, volume 59, páginas 204-217. 1976.

TURIEL, Elliot. *The development of social knowledge: Morality and convention*. Cambridge: Cambridge University Press, 1983.



■ Giselle Beiguelman, frames do vídeo generativo da série Flora Rebellis, versão 2024.
Imagens geradas com Inteligência Artificial (Style Gans)

■ [Acesse o vídeo clicando na imagem](#)





O POTENCIAL DA INTELIGÊNCIA ARTIFICIAL PARA A OTIMIZAÇÃO DO SISTEMA DE DIMENSIONAMENTO DE CONFLITOS.

DOI: <https://doi.org/10.35699/2965-6931.2023.47675>

OTÁVIO MORATO DE ANDRADE*; DIERLE NUNES**

RESUMO: Em face da demanda cada vez maior pela efetividade normativa da atuação processual-jurisdicional, os modelos de inteligência artificial (IA) têm sido implementados em diversos tribunais do Brasil, se transformando em política pública por meio da plataforma SINAPSES. Se de um lado tais sistemas são vistos como promessa para aprimorar a administração de justiça, de outro, ainda há muito a esclarecer sobre seu funcionamento, seus limites e os desafios. O presente estudo identifica três experiências de IA em tribunais brasileiros (SINAPSES; Projeto PEDRO e Projeto Victor), realizando um balanço dessas iniciativas até o momento e refletindo sobre eventuais desafios no implemento da IA no Judiciário. Conclui-se que o uso sistêmico de IA atenua vários dilemas da eficiência processual, podendo beneficiar a sociedade como um todo, desde que sua aplicação seja conciliada com parâmetros éticos e garantias fundamentais, sem prejuízo do implemento de soluções específicas de privacidade, segurança e explicabilidade.

PALAVRAS-CHAVE: inteligência artificial; judiciário; morosidade; explicabilidade. Parallels between natural and artificial intelligence from an epistemological comparison: Limitations and impacts

The potential of artificial intelligence for optimizing the conflict resolution system

ABSTRACT:

In the face of an increasingly higher demand for normative effectiveness in procedural-jurisdictional actions, artificial intelligence (AI) models have been implemented in various courts in Brazil and have already become a public policy through the SINAPSES platform. While on one hand, such systems are viewed as a promise to enhance the administration of justice, on the other hand, there is still much to clarify about their functioning, their limits, and the challenges. The present study identifies three AI experiences in Brazilian courts (SINAPSES; Project PEDRO and Project Victor), conducting an assessment of these initiatives up to the present moment and reflecting on potential challenges in implementing AI in the Judiciary. It is concluded that the systemic use of AI mitigates several dilemmas of procedural efficiency, potentially benefiting society as a whole, as long as its application is reconciled with ethical parameters and fundamental guarantees, without prejudice to the implementation of specific solutions for privacy, security, and XAI (explainable artificial intelligence).

KEYWORDS: artificial intelligence; judiciary; slowness. XAI.

* Universidade Federal de Minas Gerais/ ** Pontifícia Universidade Católica de Minas Gerais (PUCMINAS)

1 Na antiga lição de Calmon de Passos: “A efetividade por que se deve propugnar, enquanto cidadão, é a da ordem jurídica positivada, único modo pelo qual o direito nas sociedades modernas é pactuado com segurança e pode ser identificado para decidibilidade dos conflitos. Preestabelecendo esquemas e modelos, definindo previamente diretivas e limites para erradicação do arbítrio na decisão dos conflitos, institucionaliza expectativas socialmente compartilháveis que emprestam previsibilidade e segurança à convivência social civilizada. [...] Destarte, não é o processo que reclama, enquanto tal, efetividade, mas sim a tutela prometida, o ordenamento como um todo sistemático, para cuja efetividade torna-se indispensável o processo de produção do direito, não qualquer processo, e sim o devido processo constitucional. A efetividade é a marca mesma do ordenamento jurídico” (PASSOS, 1999, p. 34).

2 “...uma abordagem bottom up (de baixo para cima), que modela novos designs a partir da etiolgia dos conflitos e dos litigantes e busca criar abordagens de acesso à justiça intuitivas para os cidadãos” (NUNES, 2023).

Introdução

A efetividade normativa do sistema processual, composto pela forja normativa de alcançar acesso à justiça e devido processo democráticos, tem sido uma busca constante e premente diante da notória ineficiência do sistema/ordenamento para atingir os anseios dos cidadãos para o dimensionamento de seus conflitos¹. As dificuldades enfrentadas para garantir uma tramitação ágil e eficaz dos processos têm suscitado debates acerca dos desafios que afetam a acessibilidade e a efetividade do sistema judiciário nacional. Em 2022, o Brasil alcançou a marca de quase 80 milhões de processos judiciais em tramitação, com uma taxa de congestionamento de 76,5% no primeiro grau e 52% no segundo grau (CONJUR, 2022; CNJ, 2022).

A acumulação de processos no Judiciário possui raízes complexas, incluindo fatores como a falta de recursos humanos e técnicos e a alta litigiosidade, conhecida como “judicialização da vida” (Barroso, 2018). Muitos estudiosos também observam que essas constatações frequentemente ignoram a predominância de processos provenientes de litigantes habituais, devido ao descumprimento de direitos, e salientam a necessidade de mudar a abordagem atual de cima para baixo (*top down*) para uma perspectiva de baixo para cima (*bottom up*)² (Carvalho & Silva, 2021).

Em abordagem tradicional, Carmen Lúcia observa que há “uma óbvia dificuldade de prestação eficiente e eficaz da prestação jurisdicional constitucionalmente garantida a todos os cidadãos”. Diante do quadro crônico de inefetividade do sistema processual, o diagnóstico da Ministra é no sentido de que “a transformação do Poder Judiciário demanda uma reflexão sobre alternativas para o quadro que ora se vislumbra” (Rocha, 1998, p. 249).

As alternativas apontadas pela doutrina tradicional ao problema da inefetividade passam pelo óbvio aumento de recursos humanos e financeiros, extinção de certas modalidades recursais (desprovida de monitoramento de resultados), fomento da padronização decisória via sistema de “precedentes qualificados”, ampliação das ADRs – *Alternative Dispute Resolutions* (em português: meios alternativos de resolução de disputas) via institutos de mediação, conciliação e arbitragem, etc.

No entanto, no atual estágio há de se propugnar por novas abordagens em face da

virada tecnológica do direito processual que “estabelece um eixo interpretativo de uma justiça digital centrada no ser humano a partir do processualismo constitucional democrático, como visão macroestrutural e multidimensional” (Nunes, 2023). A virada tecnológica processual, de acordo com Dierle Nunes (2023), tem por objetivo reestruturar a forma como os litígios são mediados, partindo de uma abordagem de baixo para cima (bottom up), data-driven, que usa o grande volume de dados disponível na era digital para criar soluções customizadas aos conflitos [...] a urgência desta mudança ao considerar o número de pessoas sem acesso à justiça, a explosão de dados digitais e a necessidade de um sistema processual democrático e eficiente. Além disso, destaca a limitação das atuais legislações processuais que foram elaboradas sem considerar as especificidades da sociedade em rede e as possibilidades oferecidas pelas novas tecnologias, com a fusão aos ganhos da psicologia comportamental e ao *design*. (Nunes, 2023).

Neste contexto de busca por soluções, os modelos de inteligência artificial (IA) podem se apresentar ferramentas viáveis e promissoras para ofertar um nível de conhecimento de inúmeros aspectos do sistema processual (perfil de litígios e litigantes, índices de recorribilidade, nível de ruído das decisões), sem esquecer que permitem o dimensionamento de atividades repetitivas aliados à automação via Robotic Process Automation – RPA (em português: processo de automação robótica)³.

Diante desta conjuntura, o presente artigo tem o objetivo de investigar o funcionamento da IA e algumas contribuições que ela pode proporcionar ao Judiciário brasileiro. Para realizar tal empreitada, o trabalho está dividido em quatro etapas. Primeiramente, é formulada uma breve síntese sobre a IA, elencando seus principais conceitos e fundamentos. Em segundo lugar, apresenta-se um panorama histórico da implementação de tecnologia digital no judiciário brasileiro, que evidencia os principais marcos normativos pertinentes, dentre os quais a Lei Geral de Proteção de Dados (LGPD) e a Resolução nº 332 do CNJ (Conselho Nacional de Justiça). Em terceiro lugar, é realizado um mapeamento de algumas iniciativas em tribunais brasileiros: SINAPSES; Projeto PEDRO e Projeto Victor. A caracterização destas experiências permite compreender melhor o estado e as perspectivas da IA no Judiciário, subsidiando a discussão desenvolvida no quarto e último capítulo, que reflete acerca dos ganhos e desafios relacionados ao avanço da IA no cenário jurídico nacional.

3 “[...] mediante RPA e outras abordagens de automação se promove no Judiciário a melhoria de inúmeras atividades de gestão burocrática durante o procedimento, como, por exemplo, a realização do cumprimento de mandados de citação (projeto Mandamus do Tribunal de Justiça de Roraima), de classificação de recursos (projetos Victor do STF e Athos do STJ), agendamento de audiências, publicação de acórdãos, certifica juntada de mídias, automação das constrições judiciais (todas iniciativas do Projeto Solaria do Tribunal Regional do Trabalho da 9ª Região)” (NUNES, 2023).

No que tange à metodologia, o construto adota uma abordagem qualitativa, uma vez que a pesquisa não se propõe a quantificar o problema em análise, mas sim a realizar uma descrição e exploração detalhada do construto em questão. A intenção é descrever as características do fenômeno estudado de forma específica, contribuindo para a construção de teorias em áreas de conhecimento ainda em desenvolvimento. O raciocínio empregado será indutivo, que se diferencia do método dedutivo ao buscar inferências universais a partir da análise de fatos e dados particulares (Severino, 2013).

É importante lembrar que o avanço da IA, especialmente por meio da *deep learning*, é um fenômeno novo e com implicações ainda não totalmente esclarecidas. Nesse sentido, a principal contribuição deste estudo repousa na apresentação de iniciativas de IA ainda pouco exploradas pela literatura, que merecem maiores aprofundamentos e debates. O campo da inteligência artificial tem experimentado avanços notáveis nos últimos anos, e a expectativa é que essa área continue a progredir e se desenvolver no futuro. A apresentação de casos práticos e a identificação de desafios da IA tem o potencial de estabelecer novas conexões e estimular debates entre os pesquisadores, desempenhando um papel crucial no avanço do conhecimento.

Inteligência artificial e aprendizado de máquina

Russel & Norvig (1995) descrevem a IA como sendo a capacidade da máquina de interpretar dados de forma racional e humana, tomando decisões autônomas com base em padrões preexistentes. Para Simons (2016), é a ciência de ensinar computadores a “aprender, raciocinar, perceber, inferir, comunicar e tomar decisões como os humanos”. Outras definições incluem a “emulação de comportamento inteligente por processos computacionais” (Schalkoff, 1990) ou a “criação de máquinas com mentes reais” (Haugeland, 1985). De maneira simplificada, pode-se dizer que os modelos algorítmicos extraem correlações e padrões a partir de grandes quantidades de dados digitais, de modo a efetuar tarefas e resolver problemas (ANDRADE, 2022).

Mas como um sistema/modelo de inteligência alcança tais propósitos? Segundo o Grupo de Especialistas em IA da União Europeia (European Commission, 2019, p. 1), a”.

Esses softwares de IA também podem adaptar seus comportamentos analisando como o ambiente foi afetado por suas ações anteriores.

De forma sintética, pode-se dizer que um sistema de IA um sistema de inteligência artificial consiste em três elementos principais: sensores (entrada/input), algoritmo de processamento (lógica operacional) e atuadores (saída/output). Os sensores adquirem informações brutas do ambiente, enquanto os atuadores agem para modificar o estado desse ambiente. A essência fundamental de um sistema de IA reside na sua lógica operacional. Com base nos dados obtidos pelos sensores (entrada de dados), a lógica operacional gera resultados para os atuadores. Esses resultados podem assumir a forma de sugestões, previsões ou decisões, as quais têm o potencial de impactar o estado do ambiente (OECD, 2019). Como pontuado em outra sede:

Segundo Dierle Nunes (2023b), é possível identificar três principais fases da evolução da inteligência artificial: a IA simbólica, o aprendizado de máquina guiado por dados (machine learning) e as futuras fases, que abrangem a IA geral (não limitada a um domínio específico) e a IA superinteligente. A primeira fase engloba técnicas iniciais de IA e a abordagem simbólica, onde especialistas humanos criam regras precisas para orientar as respostas da máquina em diferentes situações. Isso facilita a compreensão do processo de decisão, identificação de erros e aprimoramentos. No entanto, essa abordagem é limitada devido à necessidade de muitas regras para representar situações complexas. A IA simbólica é aplicada em ambientes restritos, como o planejamento, onde é utilizada para criar planos de ação e otimizar o desempenho. A segunda fase, o aprendizado de máquina, permite que sistemas aprendam automaticamente a partir de dados, identificando padrões e generalizando conhecimento para uso posterior, sem programação explícita. Isso incorpora a experiência além da racionalidade.

Os ajustes feitos pelo próprio software com base na experiência do sistema, o aprendizado de máquina também pode ser aperfeiçoado por meio da intervenção humana. Desta maneira, cientistas e técnicos responsáveis pela sua implementação podem editar suas linhas de programação para corrigir erros até que o computador “aprenda” a tarefa com um nível aceitável de precisão (Miller, 2019, p. 3). Quando bem-sucedido, o aprendizado de máquina pode produzir decisões e tarefas automatizadas que se aproximam daquelas feitas por uma pessoa na mesma situação. Desta forma, o machine learning deve ser considerado como um ramo da IA, uma vez que um algoritmo eficaz é capaz de produzir resultados aparentemente “inteligentes” (Surden, 2014, p. 90).

Nunes & Andrade (2022, p. 5) explicam que existe uma “classe mais recente e avançada de algoritmos de aprendizado de máquina, valendo-se de múltiplas camadas para extrair progressivamente recursos de nível superior da entrada bruta”: o aprendizado profundo ou *deep learning* (Figura 1), que ganhou proeminência com a ampliação da capacidade computacional, desde 2010, pelo emprego das *Graphic Processing Units* – GPUs (em português: unidades de processamento gráfico). Portanto, enquanto o *machine learning* analisa dados e toma decisões com base no aprendizado, o *deep learning* cria redes neurais artificiais em camadas para aprender e tomar decisões autônomas.

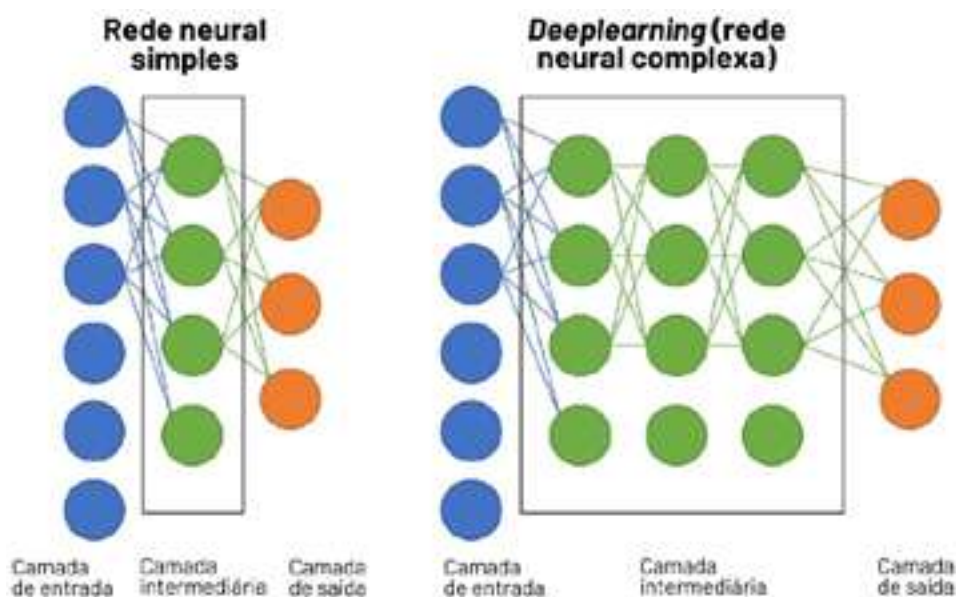


Figura 1 - Diferença entre a rede neural simples e a rede neural no deep learning

Fonte: Elaboração própria, baseado em Ryu, Chung & Choi (2021)

Lógica computacional e automatização das tarefas jurídicas

Segundo Renato Vasconcelos Magalhães (2004), a ideia de se utilizar a computação para automatizar tarefas jurídicas teve início com os trabalhos de Loevinger no final da década de 40, mas foi na década de 60 que se iniciaram as primeiras experiências práticas com o uso de algoritmos no Direito. Nas décadas de 70 e 80, como fruto

destas pesquisas, surgiram os primeiros sistemas de automação jurídica, notadamente os TAXMAN⁴ e o HYPO⁵.

Essa trajetória científica culminou no desenvolvimento de plataformas sofisticadas de inteligência artificial como o ROSS⁶ e o LawGeex⁷, softwares que passaram a ser utilizados na advocacia, sem negligenciar os recentes empregos dos grandes modelos generativos de linguagem mediante IA (do inglês: Large Generative AI Models – LGAIs), que embasam o famoso ChatGPT, que permite usos extraordinários na área jurídica, com as cautelas necessárias, em face dos riscos de alucinações máqunicas, ou seja, o fato dela inventar respostas e as apresentar com enorme confiança (STOKEL-WALKER, 2023).

Luiz Fux (2019, p. 3) relata que a International Business Machines (IBM) definiu seis categorias de potenciais aplicações da inteligência artificial ao Direito, a saber: previsão de resultados de conflitos judiciais, elaboração de peças jurídicas; revisão de contratos; identificação de padrões em decisões judiciais; rastreamento de propriedade intelectual e mecanização do faturamento de honorários.

Com esse universo de possibilidades de aplicação, surgem óbvias consequências para a vida e as profissões jurídicas, que incluem a rapidez e a melhoria dos serviços legais. Numa perspectiva mais ampla, o uso da IA no Direito tem o potencial de trazer ferramentas não apenas aos juristas, mas para toda a sociedade, na medida em que é um valioso auxílio no combate à morosidade processual e na concretização da efetividade. É o que sustentam João Alfredo Beltrão Filho e Jaqueline Cruz (2019):

A Inteligência Artificial (IA) vem sendo apontada como alternativa eficiente para reversão da morosidade tão característica do Judiciário brasileiro, visando reduzir significativamente o tempo de tramitação de processos, por meio da automação de atividades de natureza administrativa, e objetivando conferir aos magistrados um lapso temporal maior para se dedicarem mais ao pronunciamento de decisões. Assim, a utilização desta tecnologia não deverá alcançar o poder de julgamento conferido aos magistrados, e sim atividades-meio, permitindo assim que os magistrados concentrem tempo e esforços em sua atividade-fim, qual seja a de julgamento, de modo a concretizar a eficiência processual e o ideal de justiça almejado pelos cidadãos.

4 TAXMAN-I e TAXMAN-II foram softwares desenvolvidos na década de 1970, na Universidade de Rutgers (E.U.A) com o intuito de analisar taxações das reorganizações das sociedades por ações, identificando com base no caso concreto e nas leis, quais operações estariam isentas de taxação (MAGALHÃES, 2004, p. 340)

5 HYPO é um sistema projetado em 1987 para analisar casos de direito contratual estadunidense utilizando o Raciocínio Baseado em Casos. Diante de um caso apresentado pelo usuário, construiu uma apropriada argumentação com base em exemplos e precedentes (MAGALHÃES, 2004, p. 344)

6 A plataforma ROSS é capaz de responder a perguntas jurídicas específicas com base em seu banco de dados jurisprudencial e doutrinário (FUX, 2019, p. 3)

7 O sistema LawGeex foi concebido para “entender contratos de baixo risco e fazer uma revisão do contrato em poucos minutos” (FUX, 2019, p. 3)

Convém ressaltar que, no atual “estado da arte” dos softwares de automação, ainda não se produz conteúdo suficientemente preciso para uma análise totalmente automatizada. Pesquisadores apontam que a IA ainda não teria capacidade de dimensionamento ético e ponderação necessária de valores para levar em conta as especificidades de cada caso em concreto, requisitos essenciais para o pronunciamento de uma decisão (Beltrão Filho & Cruz, 2019, p. 6).

Desta forma, não se trata de delegar o poder de julgamento a estes novos sistemas computacionais, mas de utilizá-los como ferramentas de suporte e apoio aos magistrados e demais colaboradores do judiciário, de forma a reforçar a autoridade das decisões proferidas, agilizar o trâmite processual e aumentar a produtividade da Justiça, como veremos nos capítulos a seguir.

Marcos legais tecnológicos e de ia no brasil

A Lei nº 11.419/2006 estabeleceu as bases legais para a informatização do processo judicial no Brasil, estimulando o desenvolvimento de sistemas eletrônicos para a tramitação de litígios. Nessa esteira, o CNJ instituiu, notadamente através das resoluções nº 90 e 99 de 2009 e da resolução nº 185 de 2013, diretrizes para orientar a incorporação de sucessivos avanços tecnológicos no sistema judiciário.

A partir de 2010, algumas cortes passaram a utilizar o Sistema de Automação da Justiça (SAJ). A plataforma oferece ferramentas para a gestão dos autos (documental) em âmbito digital, dentre as quais o peticionamento eletrônico, a distribuição computadorizada de petições iniciais e a publicação automática com o DJe (Diário do Judiciário eletrônico). O SAJ é um marco da informatização no Poder Judiciário brasileiro. Suas funcionalidades abrangentes ajudaram a consolidar a transição do processo judicial físico para o eletrônico, facilitando tarefas, reduzindo tempos mortos e agilizando o andamento das demandas.

Entretanto, o SAJ, nem tampouco as dezenas de plataformas criadas a partir dele, sem qualquer preocupação com a interoperabilidade, têm relevância para este ensaio pelo simples fato de que são os dados digitais neles gerados e tratados que ofertam a

base para tratamento e oferta de conhecimento mediante modelos de IA. Sem a prévia etapa de digitalização, seria impensável pensar no uso de IA no Judiciário. As plataformas informatizadas de gestão e controle, que fornecem recursos para gerenciar e otimizar o trâmite processual (Krammes, 2008, p. 45), no entanto, só recentemente passaram a ser percebidas como repositórios de dados (especialmente após o DATA-JUD⁸ e o CODEX⁹) que permitem extração de conhecimento hábil à automação mais sofisticada via RPA ou para, mediante IA, dimensionar problemas e tarefas mais sofisticadas, como o apoio à decisão.

Não se pode deixar de destacar igualmente, que nos últimos anos, o avanço dos sistemas algorítmicos, da coleta e do tratamento de dados intensificou as discussões regulatórias sobre o tema. Neste cenário, foi sancionada em 2018 a Lei Geral de Proteção de Dados (LGPD), que passou a vigorar em agosto de 2020. A lei, inspirada na *General Data Protection Regulation* (GDPR) da União Europeia, tem como objetivo central a regulamentação da coleta, armazenamento, tratamento e compartilhamento de dados pessoais, visando garantir maior proteção e privacidade aos cidadãos perante as organizações brasileiras e estrangeiras que lidam com dados de indivíduos presentes no território nacional.

A LGPD introduziu conceitos antes inéditos à legislação pátria, tais como: definição ampla de dados pessoais, necessidade de consentimento do cidadão para o tratamento de seus dados, fiscalização centralizada pela Autoridade Nacional de Proteção de Dados Pessoais (ANPD¹⁰) e a imposição de penalidades substanciais em caso de não conformidade, com multas de até 2% do faturamento anual da organização no Brasil. Além disso, oferece direitos aos cidadãos, como solicitar a exclusão de dados e transferir informações para outros prestadores de serviços.

Ainda no âmbito normativo e dentro de uma tendência generalizada no Brasil de regulação na infra normatividade, sem passar pelo complexo trâmite legislativo, em agosto de 2020, o CNJ publicou a Resolução de nffl 332 um marco que estabelece princípios éticos, transparência e governança para a aplicação da IA no Poder Judiciário. Ela enfatiza a necessidade de utilizar a IA de maneira ética e igualitária, evitando discriminação e preconceitos, além de destacar a importância da transparência e da segurança de dados em modelos algorítmicos (CNJ, 2020a). Apesar do pioneirismo e da importância da norma, segundo pesquisa empírica patrocinada pela FGV, o nível

8 Base Nacional de dados do Poder Judiciário, instituído pela Resolução CNJ n. 331/2020.

9 A plataforma CODEX, instituída pela resolução nffl446/2022 do CNJ inaugura o paradigma de sistemas de processo eletrônico orientados a dados no Judiciário, sendo a principal fonte de dados estruturados e não estruturados para atender às diversas soluções tecnológicas, no âmbito do Programa Justiça 4.0. Em 2022, atingiu a marca dos 100 milhões de processos extraídos e armazenados em sua base de dados (LASALVIA, 2022). A página de acesso público ao CODEX, mantida pelo CNJ, informa em abril de 2023, que 87 tribunais brasileiros estão com seus sistemas de processos eletrônicos integrados à plataforma.

10 O tratamento de dados pessoais pelo Poder Público tem características distintas, ressaltando a necessidade de encontrar um equilíbrio entre as funções típicas do Estado e os princípios, normas e direitos definidos na LGPD. Para maiores aprofundamentos, consultar guia de tratamento de dados pessoais da ANPD: <https://www.gov.br/anpd/pt-br/documentos-e-publicacoes/documentos-de-publicacoes/guia-poder-publico-anpd-versao-final.pdf>

de *enforcement* no cumprimento das normas éticas de Resolução ainda são bastante limitadas (TAUK et al, 2023).

Em setembro de 2022 e dezembro de 2020, a Resolução nffl 335 e a Portaria nffl 271 do CNJ, respectivamente, definiram parâmetros mais específicos para a implementação da IA, visando incentivar a colaboração, transparência e divulgação de projetos entre entidades do Judiciário, além de dar diretrizes para a informatização, governança e colaboração tecnológica entre Tribunais (CNJ, 2020b; CNJ, 2020c).

Em 2023, o presidente do Senado, Rodrigo Pacheco (PSD), apresentou um projeto de lei (PL 2338/2023) que busca estabelecer o primeiro Marco Legal de Inteligência Artificial no Brasil. O projeto visa definir princípios e regras para o uso responsável da IA, assegurar a transparência dos algoritmos, proteger os direitos dos afetados e combater a discriminação. No entanto, especialistas divergem sobre os possíveis impactos: enquanto alguns veem a regulamentação como uma forma de proteção necessária, outros alertam que regulamentações rígidas podem prejudicar a inovação tecnológica e não acompanhar o rápido avanço da tecnologia. O texto, elaborado por uma comissão de juristas e especialistas, busca equilibrar a proteção dos direitos individuais com o estímulo à inovação no cenário da IA (Brasil, 2023).

Nos últimos anos, em especial durante a pandemia de Covid-19, o “Programa Justiça 4.0” vem impulsionando a modernização do Judiciário brasileiro nas seguintes frentes: a) implantação do Juízo 100% Digital; b) implantação do Balcão Virtual; desenvolvimento da Plataforma Digital do Poder Judiciário (PDPJ-Br), que amplia a automação do processo judicial eletrônico e o uso de IA; c) auxílio aos tribunais no processo de aprimoramento dos registros processuais primários e consolidação DATAJUD e d) implantação do sistema CODEX (que visa alimentar automaticamente o DATAJUD e transformar peças em texto puro, para que possam ser processadas por modelos de IA). Segundo o CNJ (2022), as inovações do “Justiça 4.0” não apenas aprimoraram a prestação jurisdicional, como também estão reduzindo os gastos do Judiciário. Até o momento, 44 tribunais (ou 67,7% das serventias judiciais) adotaram o Juízo 100% Digital, o que torna remotos e eletrônicos todos os atos processuais, incluindo audiências e julgamentos.



Algumas experiências de ia no judiciário brasileiro: caracterização de três iniciativas

11 Tribunal de Justiça do Estado de Rondônia. SINAPSES: Inteligência Artificial do TJRO desperta interesse no público da Campus Party. <https://www.tjro.jus.br/noticias/item/9729-sinapses-inteligencia-artificial-do-tjro-desperta-interesse-no-publico-da-campus-party>

12 Tribunal de Justiça do Estado de Rondônia. Inteligência Artificial desenvolvida pelo TJRO pode revolucionar o Judiciário. <https://www.tjro.jus.br/noticias/item/9472-inteligencia-artificial-desenvolvida-pelo-tjro-pode-revolucionar-o-judiciario>

13 Precedentes qualificados referem-se a decisões que estabelecem uma tese jurídica fundamental, que determina a interpretação do tribunal sobre uma questão específica a todos os casos relacionados. A relevância dos precedentes qualificados foi ressaltada pelo Código de Processo Civil (CPC) de 2015. Estes precedentes, listados no artigo 927 do CPC, ganharam força vinculativa, o que significa que devem ser seguidos pelas instâncias judiciais inferiores.

Através do “Justiça 4.0”, o CNJ (2022) revelou um aumento significativo no número de projetos de inteligência artificial no Poder Judiciário em 2022. O levantamento mapeou na plataforma SINAPSES a existência de 111 projetos em desenvolvimento nos tribunais, representando um crescimento de 171% em relação a 2021. Além disso, 53 tribunais estão atualmente envolvidos em projetos de IA, em comparação com os 32 órgãos no ano anterior. A maioria desses projetos busca melhorar a produtividade, inovar, elevar a qualidade dos serviços judiciais e reduzir custos (Rodrigues & Melo, 2020).

Nos tópicos a seguir, são apresentados três projetos de IA existentes no judiciário, que efetivamente buscam dimensionar tarefas e problemas delicados na administração da justiça. Estes sistemas consistem, em geral, na análise algorítmica de dados para otimização consistente da gestão de recursos, clusterização de procedimentos e análise de iniciativas de IA no Judiciário pátrio.

Tribunal de justiça de rondônia e cnj: “sinapses”

O Tribunal de Justiça de Rondônia foi um dos pioneiros na utilização da inteligência artificial. A partir de 2017, seus analistas utilizaram 44 mil despachos, sentenças e julgamentos no treinamento do modelo SINAPSES, com base em *machine learning*¹¹. Especialistas do TJRO foram responsáveis por otimizar essa ferramenta, implementando aprimoramentos que a tornaram adequada para utilização em diversos tribunais em todo o país. Além disso, foram realizados treinamentos para garantir o correto uso da plataforma por esses diferentes órgãos judiciários.

Os algoritmos, que foram criados e operaram inicialmente no TJRO a partir de 2018, desenvolvem, entre várias funções: i) análise computadorizada dos atos para

identificar em que fase processual se encontra a demanda, predizendo a próxima movimentação (decisão, sentença etc.); ii) pesquisa automática de jurisprudência com base na leitura computadorizada da petição inicial e iii) sugestão de palavras durante a digitação de decisões pelo magistrado, conforme o teor do texto¹².

No final de 2018, o SINAPSES foi incorporado em âmbito nacional por meio de um Acordo de Cooperação com o CNJ, figurando também, na Resolução 332/2020, como referência em solução computacional, cujo objetivo é armazenar, testar, treinar, distribuir e auditar modelos de inteligência artificial (CNJ, 2020).

A partir daí, a plataforma passou a ser provida pelo Departamento de Tecnologia da Informação do CNJ às demais entidades do Poder Judiciário, que possuem responsabilidade individual sobre a administração e supervisão dos seus próprios modelos e conjuntos de dados. Desta maneira, o SINAPSES transformou-se em um verdadeiro “hub” que conecta soluções de IA desenvolvidas em todo o Brasil, tornando-as disponíveis para adaptação às diversas rotinas dos magistrados e servidores de diferentes Tribunais, que podem personalizar os modelos de IA a partir das suas realidades e demandas.

Ao estudar o SINAPSES, Pereira & Rodrigues, (2021, p. 8) concluíram que a reunião de modelos algorítmicos pela plataforma ajudou a mitigar dois entraves para o desenvolvimento da IA no Judiciário brasileiro: a descontinuidade administrativa e a dispersão de várias iniciativas que faziam o mesmo trabalho, mas não se comunicavam entre si. Portanto, a existência de um depósito central de modelos que serve de *hub* de integração para as demais plataformas aumenta a troca de experiências e o aproveitamento de trabalho empreendido pelos técnicos de vários Tribunais.

Projeto pedro

A crescente quantidade de precedentes qualificados¹³ (superior a 2000 nos dias atuais), tem ensejado um desafio aos magistrados devido à amplitude temática e à natureza repetitiva das consultas. Não bastasse o grande volume de precedentes, sua diversidade temática também exige um esforço considerável de pesquisa, tornando o

processo de consulta cansativo, monótono e repetitivo. Diante desse quadro, pesquisadores e técnicos da UnB (Universidade de Brasília), apoiados pelo CNJ e pelo PNUD (Programa das Nações Unidas para o Desenvolvimento), desenvolveram uma plataforma que visava mitigar o problema dos precedentes por meio da IA.

Desta forma, nasceu o Projeto PEDRO (Plataforma de Extração e Descoberta de Precedentes dos Tribunais), cujo objetivo era encontrar uma solução tecnológica de apoio, não vinculante, para melhorar o ambiente de trabalho dos magistrados. Essa solução também deveria considerar a diversidade de assuntos e competências da magistratura, bem como as regras administrativas estabelecidas pelo CNJ (BALDONI & LASALVIA, 2023). O projeto está inserido sob o guarda-chuva do “Programa Justiça 4.0”, referido anteriormente.

Consoante explicação do Professor Fabiano Hartmann (2023), o foco principal do projeto foi desenvolver uma solução que permitisse a recuperação de precedentes em peças jurídicas complexas, algo que não havia sido abordado de forma abrangente na literatura acadêmica sobre IA. A solução, que baseou-se em modelos de IA não-supervisionados por meio da otimização de modelos embarcados para escalabilidade, apresenta uma interface simples para os magistrados, permitindo que insiram textos longos e recebam indicações de precedentes relevantes (apontando percentuais de similaridade de cada resultado com o *input*) para a argumentação apresentada.

A plataforma PEDRO foi entregue ao CNJ em 2023, integrando mais de 300 temas de recursos repetitivos e de repercussão geral. Os resultados obtidos com a base de dados do CNJ mostraram uma assertividade acima de 80%. De acordo com o Banco Nacional de Dados de Demandas Repetitivas, dos 1,7 milhão de processos suspensos por precedentes, mais de 1 milhão estão contemplados nos precedentes considerados no modelo (MAEJI, 2023). Além disso, especialistas pesquisadores do Direito avaliaram a solução manualmente e concluíram que ela era capaz de identificar precedentes abstratos no texto e as indicações de precedentes eram corretas para todas as petições iniciais (HARTMANN, 2023b).

Com base nisto, pode-se dizer que o Projeto PEDRO obteve êxito em desenvolver soluções baseadas em IA para auxiliar os magistrados brasileiros a lidar com a crescente quantidade e complexidade de precedentes qualificados originados de decisões em repercussão geral (STF) e repetitivos (STJ).

Supremo tribunal federal (stf): o projeto victor

O Supremo Tribunal Federal também conduziu um importante projeto de inteligência artificial, que culminou no desenvolvimento de um modelo classificador intitulado Victor, em homenagem ao ex-Ministro da Corte, Victor Nunes Leal. A utilização do robô foi concebida pelo Supremo em uma parceria com três cursos da Universidade de Brasília (UnB): Direito, Engenharia de Software e Ciência da Computação (JUNQUILHO & MAIA FILHO, 2018, p. 222).

FUX (2019, p. 11) relata que a ferramenta já é capaz de realizar quatro tipos de atividades: i) conversão de imagens em textos no processo digital; ii) separação do começo e do fim de um documento (peça processual, decisão etc.) em todo o acervo do Tribunal; separação e classificação das peças processuais mais utilizadas nas atividades do STF e iv) identificação dos temas de repercussão geral de maior incidência.

Uma das suas aplicações mais interessantes é o exame de requisitos processuais em temas de repercussão geral¹⁴. De acordo com Junquilha & Maia Filho (2018, p. 222), o robô analisa todos recursos extraordinários e agravos em recursos extraordinários que chegam à Corte, verificando se as peças cumprem o critério estabelecido pelo art. 102, § 3º¹⁵, da Constituição Federal, ou seja, se o requerente demonstrou a repercussão geral das questões constitucionais discutidas no caso.

Segundo o Ministro Dias Toffoli, essa caracterização contribui para desafogar o sistema de Justiça, pois quando uma dessas demandas é julgada, os demais processos semelhantes e a ele vinculados pelo instituto da repercussão geral podem ter o mesmo desfecho. Desta forma, além de poupar tempo para o trabalho da Justiça, a nova ferramenta pode economizar recursos humanos:

O trabalho que custaria ao servidor de um tribunal entre 40 minutos e uma hora para fazer, o software faz em cinco segundos. Nossa ideia é replicar junto aos Tribunais Regionais Federais (TRFs), aos Tribunais de Justiça, aos Tribunais Regionais do Trabalho, enfim, trata-se de uma ferramenta para toda a magistratura (TOFFOLI, 2019).

¹⁴ O instituto da repercussão geral foi criado pela Emenda Constitucional n. 45/2004 e tem por objetivo uniformizar a interpretação constitucional em casos análogos, evitando a sobrecarga da Suprema Corte. Ele cria um requisito de admissibilidade para os Recursos Extraordinários no STF. Desta forma, a questão recorrida deve ter relevância social, política, econômica ou jurídica que transcenda os interesses subjetivos da causa, gerando interesse público.

¹⁵ “Art. 102. Compete ao Supremo Tribunal Federal, precipuamente, a guarda da Constituição, cabendo-lhe: § 3º No recurso extraordinário o recorrente deverá demonstrar a repercussão geral das questões constitucionais discutidas no caso, nos termos da lei, a fim de que o Tribunal examine a admissão do recurso, somente podendo recusá-lo pela manifestação de dois terços de seus membros. (Incluída pela Emenda Constitucional nº 45, de 2004)”

¹⁶ Os Objetivos de Desenvolvimento Sustentável (ODS) constituem uma iniciativa da ONU composta por 17 metas, com prazo até 2030, que abrangem áreas cruciais como erradicação da pobreza, igualdade de gênero, saúde, educação, energia limpa, crescimento econômico inclusivo, sustentabilidade ambiental e ação climática. Os ODS visam enfrentar desafios globais complexos, promovendo a cooperação internacional e incentivando a adoção de políticas e práticas sustentáveis em todo o mundo.

Atualmente, o STF conta com uma Assessoria de Inteligência Artificial, que gerencia três modelos algorítmicos desenvolvidos para atender às demandas da corte. Além do Projeto Victor, os servidores e magistrados operam os robôs RAFA 2030 e Vitória.

O RAFA (Redes Artificiais Focadas na Agenda) 2030, visa classificar as ações judiciais de acordo com os Objetivos de Desenvolvimento Sustentável (ODS)¹⁶ da Agenda 2030 da Organização das Nações Unidas. A plataforma analisa o acervo da corte para rastrear e agrupar, através de IA, documentos como acórdãos e petições relativos aos 17 ODS. Segundo o STF (2023), o modelo algorítmico é supervisionado por análises estatísticas e OCR (*Optical Character Recognition*), e sua precisão gira em torno de 90% (MOREIRA, 2023).

A plataforma Vitória, por sua vez, está em fase final de testes. Seu objetivo é ampliar o conhecimento sobre o perfil dos processos recebidos no STF, permitindo o tratamento conjunto de temas repetidos ou similares. O sistema rastreia processos dentro do acervo do Tribunal relacionados ao mesmo tópico e os organiza de forma automática, agilizando o agrupamento de demandas e a identificação de novos temas de grande repercussão. Atualmente, Vitória tem capacidade para agregar automaticamente em torno de 5 mil processos em cerca de 2 minutos. Até então, todo esse trabalho era feito manualmente (STF, 2023).

Reflexões sobre a implementação da ia no judiciário para aprimoramento do conhecimento estratégico, gestão de atividades repetitivas e apoio à decisão

Humberto Theodoro Júnior (2004, p. 4) nos mostra que, após um século de estudos aprofundados sobre os conceitos fundamentais do Direito Processual Civil, os especialistas perceberam que a sociedade desejava uma prestação jurisdicional mais eficiente e acessível. Desta forma, vem se buscando progressivamente uma proteção adequada e legítima dos direitos individuais, sensível a diferentes conflitos e camadas

sociais.

Neste sentido, a garantia de acesso à Justiça e a efetividade normativa prestada, como destacado inicialmente, se consolidaram como demandas primordiais da sociedade contemporânea agora em conformidade com a virada tecnológica do direito processual.

Portanto, o Poder Judiciário tem sido cada vez mais cobrado pela sociedade, no sentido de oferecer respostas adequadas e adaptadas aos novos conhecimentos e abordagens ofertadas pelas perspectivas atinentes ao design de sistema de disputas (DSD) permitindo uma customização de soluções embasadas em dados e desde os gatilhos das litigâncias. Todavia, a ausência de conhecimento dos verdadeiros gatilhos das ações, as restrições orçamentárias e os limitados recursos humanos se revelam insuficientes para dar conta do gigantesco volume de processos (Beltrão Filho & Cruz, 2019). Um dos grandes entraves enfrentados pelo Judiciário nacional tem sido a conciliação desse clamor social – totalmente legítimo, diga-se de passagem – por uma justiça centrada no ser humano e no dimensionamento de suas necessidades (Nunes, 2023), com os gargalos orçamentários existentes no Estado brasileiro, sobretudo em tempos de estagnação econômica.

Em face desse cenário desafiador, alguns Tribunais brasileiros passaram a investir em soluções tecnológicas para combater seus dilemas, mas ainda falta uma mudança consistente que leve em consideração todos os afetados (*stakeholders*) de modo a se implementar uma efetiva postura *bottom up*, mais hábil à complexidade dos problemas a se dimensionar (Nunes, 2023) Essas soluções incluem desde a simples informatização de procedimentos que antes eram físicos (como a digitalização de processos), passando pela automação por RPA – *Robotic Process Automation* (em português: automação robótica de processos), até o uso de modelos de IA, como veremos adiante. Tal abordagem deve levar a sério a virada tecnológica:

A virada tecnológica do direito processual, como eixo interpretativo de uma justiça digital centrada no ser humano, busca mediante uma fusão de conhecimentos e novas tecnologias permite uma abordagem bottom-up dos conflitos e dos litigantes permitindo a construção de designs hábeis às necessidades dos cidadãos e que possa ser a

17 A explicabilidade da IA consiste em transformar os algoritmos complexos e opacos (“caixas-pretas”) em “caixas de vidro” nas quais prevalece a transparência, de modo que os modelos computacionais possam, de forma ativa, explicar e justificar seus processos de raciocínio. Acredita-se que a explicabilidade pode ajudar a reconhecer vieses, apurar responsabilidades e tornar os sistemas mais confiáveis e seguros (NUNES & ANDRADE, 2022, p. 23).

18 Segundo Nunes (2018, p. 6), um robô que avalia o risco de reincidência de criminosos nos Estados Unidos passou a classificar erroneamente os acusados negros como prováveis reincidentes, e os acusados brancos como indivíduos de baixo risco.

19 Nunes (2018, p. 8) relata que “(...) os algoritmos utilizados nas ferramentas de inteligência artificial são obscuros para a maior parte da população – algumas vezes até para seus programadores (...) – o que os torna, de certa forma, inatácáveis.”

20 Um relatório do Governo Norueguês, “Artificial intelligence and privacy” (2018), descreve os riscos e preocupações governamentais relacionados ao armazenamento e manipulação de dados pessoais por máquinas. Disponível em: <https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf>

mais intuitiva, convidativa e adequada para os conflitos, numa revisão da percepção de um modelo multiportas. Mediante uma perspectiva data-driven se permite uma compreensão precisa e dinâmica do ingresso dos conflitos e um gerenciamento preventivo e ativo que promove a possibilidade de acompanhamento de toda a jornada do dimensionamento dos conflitos, desde a etapa prévia ao ajuizamento de processos, até sua verificação ao longo de todo o iter procedimental, viabilizando o acompanhamento dos gargalos no procedimento, a análise da performance e com tais conhecimentos a possibilidade de criação de interfaces (procedimentos) com usabilidade aprimorada que permitiriam reduzir até mesmo as assimetrias informacionais com designs centrados no ser humano, emprego de legal design e plain language, e oferta de soluções mais relacionadas e adaptadas às necessidades (NUNES, 2023)

Na medida em que governos e empresas implantam tecnologias de inteligência artificial, surge a preocupação em compatibilizar esses avanços científicos com os direitos fundamentais, com vistas a fornecer uma justiça centrada no ser humano. Isso porque o uso de modelos computacionais, como toda revolução tecnológica, oferecerá certos riscos e dilemas, que devem ser tratados de forma adequada e proporcional, preservando a ética e os direitos fundamentais.

Considerando que os direitos fundamentais são fundantes da sociedade e devem se contrastar a qualquer uso indevido de poderes, inclusive os agora algorítmicos, ofertando um ambiente jurídico corretivo (contrafático), faz-se primordial harmonizar o uso da IA com tais preceitos e se fundar novas bases para um devido processo constitucional tecnológico. Afinal, se o uso da inteligência artificial tem o intuito de aperfeiçoar o bem-estar e a liberdade, seu grande beneficiário é, pois, o próprio indivíduo, razão pela qual a aplicação da IA jamais poderia se distanciar dos princípios éticos de uma sociedade ou confrontar as garantias fundamentais franquadas pelo Estado Democrático de Direito.

Neste sentido, a União Europeia divulgou as “Diretrizes Éticas para a Inteligência Artificial Confiável”. O documento define orientações para o desenvolvimento de tecnologias computacionais cognitivas, justamente com base nos erros, riscos e impactos negativos que o raciocínio automatizado pode produzir (Frazão, 2019). Tais orientações estão sustentadas em dois pilares básicos: direitos fundamentais (dignidade humana, liberdade do indivíduo, igualdade, solidariedade, não-discriminação, respeito

aos cidadãos e ao Estado de Democrático Direito) (“Diretrizes éticas para a I.A. Confiável”, 2019, p. 9) e os princípios éticos do respeito à autonomia humana, prevenção de danos, justiça e *explicabilidade*¹⁷ (idem, p. 10).

Com base nessas diretrizes éticas, o guia (idem, p. 14) também emitiu sete orientações técnicas que devem ser seguidas por governos e corporações no desenvolvimento e utilização da inteligência artificial. São elas: i) fiscalização; ii) segurança técnica; iii) privacidade e governança de dados; iv) transparência; v) diversidade, não-discriminação e equidade; vi) bem-estar social e ambiental; vii) responsabilização.

Assim como a União Europeia, outros países, como o Brasil têm se proposto a traçar critérios para o emprego de inteligência artificial. Estes novos marcos regulatórios, portanto, têm o intuito de oferecer normas preventivas e estipular soluções para os impactos gerados pelo avanço da computação cognitiva.

Entre as inúmeras implicações da IA descritas na doutrina, podemos elencar: a) presença de vieses cognitivos nos algoritmos de IA, que eventualmente produzem resultados discriminatórios¹⁸ ou pouco transparentes¹⁹ (Nunes, 2018, p. 6-8); b) as ameaças à privacidade, na medida em que os robôs podem acessar e interpretar informações pessoais do indivíduo²⁰; c) a possível substituição dos humanos pelas máquinas no ambiente de trabalho (Unger, p. 68) e (Acemoglu & Restrepo, 2018, p. 3); d) o problema da responsabilização no caso do cometimento de ilícito por uma máquina (FUX, 2019, p. 5); e) o uso militar de armamentos robóticos altamente destrutivos²⁰; f) a hipótese das consequências não-intencionais, segundo a qual os sistemas computacionais sofreriam, uma explosão de inteligência, agregando um nível de complexidade e sofisticação incontrolável pelos seres humanos.

Ante a todas essas implicações e desafios, enxergar a inteligência artificial à luz da ética e dos direitos fundamentais é uma janela de oportunidade para moldar seu desenvolvimento, garantindo que as Tecnologias de informação e comunicação (TICs), com destaque às Inteligências Artificiais, se desenvolvam de maneira segura, oferecendo um ambiente de operação confiável para seus próprios usuários. Por isso, o Estado e o Direito devem extrair da IA seus incontestáveis benefícios sem, no entanto, se furtar de regulamentar o uso e o desenvolvimento dessas novas tecnologias à luz da ética e dos princípios insculpidos na Constituição.

20 O Professor de Oxford, Nick Bostrom afirma que a inteligência artificial gera os chamados riscos existenciais: “ameaças que podem causar nossa extinção ou destruir o potencial da vida inteligente que origina a Terra” (tradução nossa). Disponível em <https://nickbostrom.com/existential/risks.html>

Considerações finais

Após apresentar conceitos fundamentais da IA e contextualizar os marcos regulatórios desta tecnologia no Brasil, este trabalho apresentou três experiências práticas de implementação de modelos algorítmicos em Tribunais (SINAPSES, Projeto PEDRO e Projeto Victor), que subsidiaram reflexões sobre a conciliação técnica de sua implementação com ditames éticos e normativos.

Com efeito, é preocupante que as análises e o uso das tecnologias se limitem a tentar dimensionar abordagens típicas de contextos prévios à explosão de dados digitais que, uma vez tratados, podem ofertar conhecimento hábil à verdadeira percepção de um diagnóstico holístico dos problemas dos cidadãos e das virtudes e dilemas do enorme aparato jurisdicional delineado para atendê-los, somado a iniciativas extraestatais, que, por vezes, apresenta um olhar distorcido e ligado a velhas práticas e percepções das litigâncias e de como dimensioná-las e da real eficiência dos atuais meios de dimensionamento dos conflitos.

Para além da automatização informática, que dá eficiência ao armazenamento, integração e acesso de dados, existe a possibilidade de se utilizarem modelos de inteligência artificial no suporte à gestão e, com um uso muito cuidadoso, de apoio às decisões judiciais. Estes sistemas, podem “agilizar a transmissão das informações, expondo analogias e contradições que seriam difíceis de identificar se os dados fossem analisados de forma manual ou separadamente” (Junquilho & Maia Filho, 2018, p. 231).

Convém ressaltar, no entanto, que não se trata de delegar à máquina o poder de julgamento, até porque isso na atualidade é inviável, mas de utilizá-la como ferramenta para aprimorar o trâmite processual e aperfeiçoar a tomada de decisões pelos magistrados, na medida em que a IA auxilia a checagem de dados, viabiliza a indexação de informações, triangulação de jurisprudência mediante pesquisas semânticas e fornece um poderoso auxílio para a automatização de tarefas repetitivas.

As iniciativas pioneiras no TJRO, Supremo Tribunal Federal, Superior Tribunal de Justiça e em outras cortes judiciais do país têm evidenciado que técnicas de IA são capazes de reduzir significativamente o tempo de tramitação de processos e ofertar um ambiente hábil a uma justiça centrada no ser humano.

Como toda virada tecnológica, a inteligência artificial oferece novos riscos e dilemas morais. Esses desafios, que incluem questões de privacidade, transparência e segurança de dados, devem ser tratados à luz dos valores éticos e de uma normatização sofisticada, que promova uma regulação embasada em riscos e direitos, que não obste a inovação, mas que prioritariamente permita aos cidadãos o auferimento de meios hábeis e usáveis de dimensionamento de conflitos embasados no devido processo e, em última análise, numa perspectiva jurídica mais adaptada ao Estado Democrático de Direito.

Referências

- ACEMOGLU, Daron; RESTREPO, Pascual. *Artificial Intelligence, Automation and Work*. NBER, Working Paper n. 24196, Janeiro de 2018.
- ANDRADE, Otávio Morato de. *Governamentalidade algorítmica: democracia em risco?* 1^{ff} ed. São Paulo: Dialética, 2022. 224 p.
- BALDONI, Marina; LASALVIA, Raquel. *Justiça 4.0 apresenta projetos de uso de IA desenvolvidos por universidades parceiras*. Website do CNJ. 23 de junho de 2023.
- BARROSO, Luís Roberto. *A judicialização da vida e o papel do Supremo Tribunal Federal*. Ed. Fórum. Brasília, 2018.
- BELTRÃO FILHO, João Alfredo; CRUZ, Jaqueline Keila Leite. *O uso da inteligência artificial na análise de processos no Judiciário como instrumento de concretização da eficiência processual*. CPC nos Tribunais. *Rev Jota*. Mai 2019.
- BRASIL. *Projeto de Lei nº 2338/2023, de autoria do Senador Rodrigo Pacheco*. 3 de maio de 2023.
- CONJUR. *Anuário de Justiça Brasil 2022*. Consultor Jurídico, 2022.
- CNJ (CONSELHO NACIONAL DE JUSTIÇA). *Relatório Justiça em Números*, 2022.
- CNJ (CONSELHO NACIONAL DE JUSTIÇA). *Resolução nº 332*. 21 de agosto de 2020.
- CNJ (CONSELHO NACIONAL DE JUSTIÇA). *Resolução nº 335*. 29 de setembro de 2020(b).
- CNJ (CONSELHO NACIONAL DE JUSTIÇA). *Portaria nº 271*. 4 de dezembro de 2020(c).
- EUROPEAN COMMISSION. *High-Level Expert Group on Artificial Intelligence. A definition of Artificial Intelligence: main capabilities and scientific disciplines*. Abril de 2019.
- EUROPEAN COMMISSION. *High-Level Expert Group on Artificial Intelligence. Ethics guidelines for trustworthy AI (Diretrizes éticas para uma inteligência artificial confiável)*. Abril de 2019(b).
- FRAZÃO, Ana. *Quais devem ser os parâmetros éticos e jurídicos para a utilização da IA?* *Rev. Jota*, 24 de abril de 2019.
- FUX, Luiz. *Palestra sobre Inteligência artificial*. *Conjur*. Fev 2019. Acesso em: 8 jun. de 2019.
- HAUGELAND, John. *Artificial Intelligence: The Very Idea*. Cambridge: MIT Press, Massachusetts, 1985.
- HARTMANN, Fabiano. *Apresentação sobre precedentes qualificados*. *Diretório de Inteligência Artificial da UnB (DR.IA)*. 2023.
- HARTMANN, Fabiano. *Relatório sobre precedentes qualificados*. Documento interno cedido pelo autor. *Diretório de Inteligência Artificial da UnB (DR.IA)*. 2023(b).
- HELM, Louie & MUEHLHAUSER, Luke. *Intelligence Explosion and Machine Ethics*. 2012.
- JUNQUILHO, Tainá; MAIA FILHO Mamede Said. *Projeto Victor: perspectivas de aplicação da inteligência artificial ao direito*. *Revista de Direitos e Garantias Fundamentais*, Vitória, v. 19, n. 3, p. 219-238, set./dez. 2018.

KRAMMES, Alexandre Golin. **Aplicação de workflow em processos judiciais eletrônicos**. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento. Florianópolis, 2008.

LASALVIA, Raquel. **Codex atinge mais de 100 milhões de processos judiciais armazenados**. **Website do CNJ**. 22 de novembro de 2022.

MAEJI, Vanessa. **Parceria com a UnB desenvolve modelo de IA de verificação automática de precedentes qualificados**. **Website do CNJ**. 24 de julho de 2023.

MILLER, Sterling. **Artificial intelligence and its impact on legal technology: to boldly go where no legal department has gone before**. Thomson Reuters, So45388, P1/4-17, 2017.

MOREIRA, Ariene. **RAFA: uso de IA pelo STF abrange Agenda 2030. Práticas ESG**. Acesso em: 4 abr. de 2023.

NUNES, Dierle José Coelho; ANDRADE, Otávio Morato de. **O uso da inteligência artificial explicável enquanto ferramenta para compreender decisões automatizadas: possível caminho para aumentar a legitimidade e confiabilidade dos modelos algorítmicos?** *Revista Eletrônica do Curso de Direito da UFSM*, v. 18(1), p. e69329, Santa Maria, 2023.

NUNES, Dierle José Coelho. **Virada tecnológica no direito processual: fusão de conhecimentos para geração de uma nova justiça centrada no ser humano**. *Revista de Processo*, São Paulo, 2023. No prelo.

NUNES, Dierle José Coelho. **Uma introdução sobre o uso das Inteligências Artificiais analíticas e generativas no Direito Processual** In VADELL, Lorenzo; JUNOY, Joan; CAMPOS, Santiago; BUENO, Cassio; MITIDIERO, Daniel; DOTTI, Rogéria; Hill, Flávia. **El sistema procesal del siglo XXI: novos retos**. *Anais das XXVII Jornadas Iberoamericanas de Derecho Procesal*. Londrina: Thoth, 2023b.

OECD. **Artificial Intelligence in Society**. OECD Publishing, Paris, 2019. Disponível em: <https://doi.org/10.1787/eedfee77-en>. Acesso em: 5 ago. de 2023.

PASSOS, José Joaquim Calmon de. **Cidadania e efetividade do processo**. *Revista Síntese de Direito Civil e Processual Civil*, n. 1, Set-Out 1999.

PEREIRA, João Carlos Murta; RODRIGUES, Marcos Vinícius Jardim Rodrigues. **A Plataforma SINAPSES e a Continuidade dos Modelos de IA no Judiciário**. *Anais do ENAJUS* 2021.

ROCHA, Cármen Lúcia Antunes. **A reforma do poder judiciário**. *Revista de Informação Legislativa*, v. 35, n. 137, p. 239-254, jan./mar. 1998

RODRIGUES, Alex; MELO, Jeferson. **Sinapses: plataforma de inteligência artificial conquista prêmio na Expojud**. **Website do CNJ**. 14 de outubro de 2020. Disponível: <https://www.cnj.jus.br/sinapses-plataforma-de-inteligencia-artificial-conquista-premio-na-expojud/>

RUSSEL, Stuart. & NORVIG, Peter. **Artificial Intelligence: a modern approach**. Prentice-Hall, Inc. Simon & Schuster Company Englewood Cliffs, 1995, New Jersey.

RYU, Jeong Yeop; CHUNG, Ho Yun; CHOI, Kang Young. **Potential role of artificial intelligence in craniofacial surgery**. *Archives of Craniofacial Surgery* 22(5):223-231. Outubro de 2021.

SCHALKOFF, Robert. **Artificial Intelligence: An Engineering Approach**. Ed. McGraw-Hill, 1990.

SILVA, Paulo; CARVALHO, Natália. **O “Grande Jogador” - Como atua o Judiciário na administração**

da litigiosidade repetitiva. *Revista de Direito Brasileira*, v. 28, n. 11, p. 300-321, Jan./Abr. 2021.

SIMONS, John. *Tomorrow's Business Leaders Learn How to Work with A.I.* *The Wall Street Journal*. Novembro de 2016.

STF (SUPREMO TRIBUNAL FEDERAL). *STF amplia emprego de Inteligência Artificial*. Assessoria de Imprensa do STF. 9 de junho de 2023. Disponível: <https://portal.stf.jus.br/noticias/verNoticiaDetalhe.asp?idConteudo=508710&ori=1>. Acesso em 14 de agosto de 2023.

STOKEL-WALKER, Cris. *Generative AI Is Coming For the Lawyers*. *Wired*. 21 de fevereiro de 2023. Disponível em: <https://www.wired.co.uk/article/generative-ai-is-coming-for-the-lawyers>. Acesso em: 13 ago. de 2023.

SURDEN, Harry. *Machine learning and law*. *Washington Law Review*, 89 pp. 87-115, 2014

TAUK, Caroline; NUNES, Dierle José Coelho; BRAGANÇA, Fernanda; COELHO, José Leovigildo; LOSS, Juliana; BRAGA, Renata. *Inteligência artificial: tecnologia aplicada à gestão dos conflitos no âmbito do poder judiciário brasileiro*. Relatório. FGV Conhecimento, 2023.

THEODORO JUNIOR, Humberto. *Celeridade e efetividade da prestação jurisdicional. insuficiência da reforma das leis processuais*. *Rev. Academia Processual de Direito Civil*. Jun 2004.

TOFFOLI, Dias. *Judiciário deve julgar o passado, deixando que a política conduza o presente e o futuro*. *Rev. Consultor Jurídico*, Maio de 2019.

UNGER, Roberto Mangabeira. *Depois do colonialismo mental: repensar e reorganizar o Brasil*. Autonomia literária, 2018.



■ Giselle Beiguelman, frames do vídeo generativo da série Flora Rebellis, versão 2024.
Imagens geradas com Inteligência Artificial (Style Gans)

■ [Acesse o vídeo clicando na imagem](#)





ÉTICA E INTELIGÊNCIA ARTIFICIAL: Desafios e melhores práticas

DOI: <https://doi.org/10.35699/2965-6931.2023.47673>

CAROLINA DE MELO NUNES LOPES*; JÚLIA CASTRO MENDES**

RESUMO: O rápido avanço da Inteligência Artificial (IA) trouxe consigo a necessidade de compreensão de seu impacto social e implicações éticas. Nesse sentido, este artigo levantou e discutiu as principais questões éticas relacionadas à IA, os principais obstáculos atuais no desenvolvimento de algoritmos de aprendizado de máquina, e as melhores práticas para desenvolver algoritmos éticos e justos. Os vieses podem se perpetuar facilmente através do uso de dados desbalanceados e de correlações infundadas. Diante disso, a colaboração entre desenvolvedores de algoritmos e outros especialistas torna-se fundamental para compreender diversas perspectivas e identificar as sutis formas de propagação dos preconceitos. A ética dos algoritmos não é uma questão que será solucionada apenas através de uma abordagem tecnológica - essa temática envolve também assuntos de ordem social, cultural, jurídica e política. Portanto, o desenvolvimento tecnológico e a responsabilidade social devem caminhar lado a lado, a fim de evitar o agravamento das diferenças sociais.

PALAVRAS-CHAVE: inteligência artificial; ética; aprendizado de máquina; aprendizado profundo; viés algoritmo

Ethics and artificial intelligence: Challenges and best practices

ABSTRACT: The rapid advancement of artificial intelligence (AI) brought with it the need to understand its social impact and ethical implications. In this sense, this article raised and discussed the main ethical issues related to AI, the current primary obstacles in the development of machine learning algorithms, and the best practices to develop ethical and fair algorithms. Biases can easily perpetuate themselves through the use of unbalanced datasets and unsubstantiated correlations. Therefore, collaboration between algorithm developers and other specialists becomes essential to understand different perspectives and identify the subtle forms of the propagation of prejudices. The ethics of algorithms is not an issue that will be solved only through a technological approach - this theme also involves social, cultural, legal, and political aspects. Therefore, technological development and social responsibility must go hand in hand to avoid the aggravation of social differences.

KEYWORDS: artificial intelligence; ethics; machine learning; deep learning; algorithm bias.

* Universidade Federal de Ouro Preto / ** Universidade Federal de Juiz de Fora

Introdução

A quarta revolução industrial impulsionou a aplicação de técnicas de Inteligência Artificial (IA) em diversos setores da sociedade. Muitas vezes não nos damos conta que interagimos com sistemas inteligentes o tempo todo, seja na sugestão de palavras do corretor ortográfico, ou nos anúncios das redes sociais (Garcia, 2020). Cada vez mais, os algoritmos mediam processos, transações bancárias, decisões governamentais e interferem na forma como nós nos percebemos e interagimos socialmente (Mittelstadd *al.*, 2016).

É inegável que a transformação digital e as aplicações proporcionadas pela IA trouxeram inúmeros benefícios para diversos setores da sociedade (Winfield & Jirotko, 2018; Doneda *et al.*, 2018). No entanto, o desenvolvimento dessas tecnologias precisa vir acompanhado de reflexões éticas e morais (Rosetti; Angeluci, 2021). Como exemplos, o uso indevido de dados e informações pessoais, os vieses propagados pelos algoritmos, e a própria ética das máquinas, como nos carros autônomos, são questões que surgiram com o avanço da Inteligência Artificial nos últimos anos (Winfield; Jirotko, 2018).

Um dos principais dilemas éticos reside no viés algoritmo. Os algoritmos de aprendizado de máquina são treinados a partir de um conjunto de dados (Garcia, 2020). Se esses dados estiverem impregnados de vieses de gênero, raça, ou qualquer outro, o algoritmo irá perpetuar esses vieses (Garcia, 2020). Em 2018, por exemplo, a empresa *Amazon* contou com um sistema inteligente para fazer uma pré-seleção de currículos em um processo seletivo (Garcia, 2020). Para treinar esse sistema, a empresa utilizou a base de dados dos próprios funcionários, a fim de que o algoritmo selecionasse pessoas com características semelhantes às que já trabalhavam na empresa. O que não foi levado em conta pelos desenvolvedores é que a base de dados era majoritariamente masculina. Dessa forma, nenhuma mulher foi selecionada pelo algoritmo (Garcia, 2020).

Além disso, a transparência e “explicabilidade” (ou interpretabilidade) dos sistemas de IA são desafiadoras, uma vez que muitos modelos de aprendizado de máquinas são como “caixas pretas”, ou seja, não permitem o acompanhamento das etapas intermediárias, ou dos cálculos que levam a um dado resultado ou decisão (Mittelstadt

et al., 2016). As lacunas entre o design, a operação de algoritmos e a nossa compreensão de suas implicações éticas podem ter consequências graves que afetam indivíduos, bem como toda a sociedade (Mittelstadt *et al.*, 2016). Para Garcia (2020), mais grave ainda é que as respostas geradas por esses algoritmos estão revestidas de mérito por serem resultado de testes estatísticos e matemáticos, o que confere a eles uma falsa imparcialidade.

Diante deste contexto, este estudo tem como objetivo levantar e discutir as questões éticas predominantes associadas à aplicação de tecnologias de IA, bem como relacionar as melhores práticas a serem adotadas neste contexto. Para tanto, foi realizada uma pesquisa bibliográfica de natureza exploratória, não-sistemática, utilizando as palavras-chave: Ética; Inteligência Artificial e Aprendizado de máquina, na ferramenta de pesquisa convencional do Google, no Google Acadêmico, Scopus e Scielo, no período de março a maio de 2023. Assim, a seguir, serão discutidas as principais preocupações em relação à ética dos algoritmos, alguns obstáculos do aprendizado de máquinas, do aprendizado profundo e as melhores práticas a serem adotadas para mitigar essas questões em projetos futuros, de qualquer área, envolvendo IA.

A ética dos algoritmos

Determinar o impacto ético de um algoritmo é um trabalho árduo (Mittelstadt *et al.*, 2016). Segundo Rosseti e Angeluci (2021), identificar a influência da subjetividade humana no projeto e configuração de algoritmos geralmente requer uma investigação multidisciplinar e de longo prazo, uma vez que a incerteza e a opacidade sobre como funcionam alguns algoritmos dificultam o seu entendimento (Winfield; Jirotko, 2018). Para Mittelstadt *et al.* (2016), as principais preocupações relacionadas à ética dos algoritmos são: evidências inconclusivas; evidências incompreensíveis; evidências equivocadas; resultados injustos; privacidade e rastreabilidade.

1.2 EVIDÊNCIA INCONCLUSIVAS E AÇÕES INJUSTIFICADAS

Os algoritmos de decisão e a mineração de dados buscam estabelecer padrões e correlações em grandes conjuntos de dados (Mittelstadt *et al.*, 2016). No entanto, nem toda correlação encontrada pelos algoritmos tem uma relação de causa e efeito, o que

pode gerar ações injustificadas (Doneda *et al.*, 2018; Rossetti; Angeluci, 2021). Assim, a decisão determinada por um algoritmo pode se basear em uma correlação estatística que não representa necessariamente uma relação causal, sem qualquer relação com o assunto objeto da decisão (Doneda *et al.*, 2018).

Nesse sentido, destaca-se a polêmica envolvendo a empresa alemã, Schufa, que presta serviços de proteção ao crédito (Doneda *et al.*, 2018). A empresa, ao avaliar o risco de inadimplência, classificava como critério negativo o pedido de consumidores para acessar seus próprios dados. Ou seja, a empresa penalizava consumidores que queriam contratar crédito simplesmente pelo exercício de um direito do consumidor. Existia aí uma relação de correlação (é mais provável que pessoas que tenham dificuldades financeiras queiram consultar seu próprio *score* de crédito), mas não de causa (essa consulta não é uma razão das dificuldades financeiras ou da inadimplência). Dessa forma, a Alemanha vedou essa prática na reforma da Lei Federal de Proteção de Dados (LGPD), de 2009 (Doneda *et al.*, 2018).

1.3 EVIDÊNCIAS INCOMPREENSÍVEIS E OPACIDADE

Quando os dados são processados para produzir evidências para uma conclusão, é esperado que exista uma relação compreensível e acessível entre esses dados e as conclusões obtidas (Mittelstadt *et al.*, 2016; Winfield; Jirotko, 2018; Rossetti; Angeluci, 2021). No entanto, muitos algoritmos de aprendizado de máquina, principalmente de aprendizagem profunda (*deep-learning*), atuam como uma caixa-preta - ou seja, é possível visualizar os dados de entrada e a saída, mas o funcionamento interno é obscuro e não facilmente interpretável pelos humanos (Winfield; Jirotko, 2018).

A opacidade dos algoritmos pode ser problemática em várias áreas, especialmente em setores críticos, como saúde, justiça, transporte e finanças. Quando um algoritmo de aprendizado de máquina toma uma decisão importante, como aprovar um empréstimo, sugerir tratamentos médicos, selecionar dosagens otimizadas de fármacos, dimensionar estruturas civis ou até mesmo guiar um carro autônomo, é essencial que haja uma explicação compreensível sobre como a decisão foi tomada. Isso é especialmente relevante quando há potenciais consequências negativas, como a integridade física, a saúde mental e o bem-estar de pessoas.

Alguns programadores alegam que a opacidade se deve à necessidade de proteger seus trabalhos da concorrência e à privacidade dos dados - o que cria um conflito ético

entre os princípios da transparência e da privacidade (Rossetti; Angeluci, 2021). No entanto, para que os algoritmos de decisão sejam bem aceitos pela sociedade, entender seu funcionamento é essencial (Winfield; Jirotko, 2018). Por essa razão, a interpretabilidade e a transparência de algoritmos têm sido focos crescentes de pesquisa em IA. No caso dos carros autônomos, por exemplo, que já geraram acidentes fatais (Stilgoe, 2018), há, claramente, uma necessidade urgente de transparência para descobrir como e por que esses acidentes ocorreram (Winfield; Jirotko, 2018).

Evidências equivocadas e os vieses

A automação das tomadas de decisões pelos algoritmos muitas vezes é justificada por uma suposta ausência de viés nesses processos, diferente das decisões humanas (Mittelstadt *et al.*, 2016). Essa crença é insustentável, conforme relatam estudos que demonstram que os algoritmos, inevitavelmente, tomam decisões tendenciosas (Doneda *et al.*, 2018; Garcia, 2020; Karale, 2021; Rossetti e Angeluci, 2021). A qualidade das decisões automatizadas baseadas em algoritmos está intimamente relacionada com a qualidade dos dados processados (Garcia, 2020; Mittelstadt *et al.*, 2016; Rossetti e Angeluci, 2021). Dessa forma, se o algoritmo é desenvolvido a partir de dados preconceituosos, ele reproduz, de forma automatizada, os mesmos padrões preconceituosos (Doneda *et al.*, 2018).

A Inteligência Artificial tem o potencial de revolucionar a área da saúde, incluindo a detecção de câncer de pele (Adamson; Smith, 2018). No entanto, se não for desenvolvida de forma inclusiva, a IA pode agravar as diferenças de gênero, raça, entre outras. Segundo Adamson e Smith (2018), na área da dermatologia, a maioria dos programas de detecção de doenças de pele é treinado com dados de pele clara, o que pode resultar em resultados enviesados e menor desempenho em imagens de pele mais escura. Essa limitação pode ter consequências preocupantes no diagnóstico de melanoma, que pode se apresentar de forma diferente em peles mais escuras (Adamson; Smith, 2018). De acordo com os autores, para solucionar essa questão, os algoritmos devem ser treinados para reconhecer o melanoma em todos os tipos de pele, e os repositórios devem

incluir mais fotos com condições de pele diversificadas. Caso contrário, as tecnologias de aprendizado de máquina e aprendizado profundo correm o risco de agravar ainda mais as disparidades na área da saúde.

Outro exemplo do viés algoritmo é o caso do *Google Fotos*, que em 2015 etiquetava pessoas negras como gorilas (Rossetti; Angeluci, 2021). A diversidade e a representatividade dos dados são fundamentais para evitar a reprodução de preconceitos e desigualdades no desenvolvimento de tecnologias de IA. Isso é essencial para diversas áreas que usam imagens de seres humanos além do diagnóstico de doenças, como as de monitoramento de segurança em canteiros de obras, identificação de criminosos, classificação de massa corporal, leitura de linguagem de sinais, reconhecimento de gestos, entre outros. Assim, a busca por modelos mais justos e equitativos é um dos maiores desafios no desenvolvimento de algoritmos de aprendizado de máquina éticos e confiáveis.

Resultados injustos e discriminação

Decisões automatizadas que contenham vieses podem levar à discriminação (Rossetti; Angeluci, 2021). O viés está relacionado à própria formulação da decisão, enquanto a discriminação refere-se aos resultados negativos desproporcionais resultantes das decisões algorítmicas (Mittelstadt; Allo, *et al.*, 2016). Em outras palavras, o viés está presente na forma como o algoritmo toma decisões, e a discriminação é o efeito prejudicial que essas decisões podem ter em termos de impacto desigual.

Obermeyer *et al.* (2019) publicaram um estudo mostrando o viés racial em sistemas da área da saúde nos Estados Unidos. Segundo o estudo, uma empresa de seguro saúde norte-americana, com objetivo de reduzir seus custos, decidiu oferecer tratamentos preventivos para pacientes com doenças crônicas em estado grave. Para selecionar os casos mais graves, a empresa considerou o número de vezes que o paciente utilizava o sistema de saúde. Quanto mais o paciente usava o sistema, mais crítico era considerado o seu caso. Para evitar vieses, dados como raça e gênero foram excluídos

da base de dados do algoritmo. No entanto, a grande maioria dos selecionados foram pessoas brancas, o que fez os pesquisadores se perguntarem: se a base de dados não continha informações sobre raça, por que o algoritmo ainda privilegiava os brancos? O que se conclui é que pessoas negras utilizavam menos o sistema de saúde por não poderem arcar com os custos de coparticipação. Dessa forma, o conjunto de dados não continha muitas informações de negros que utilizavam o sistema de saúde, gerando um algoritmo racista.

A mera exclusão direta de características como raça e gênero da base de dados não é suficiente para garantir a equidade das decisões algorítmicas (Obermeyer *et al.*, 2019). Portanto, a conscientização sobre a interação crítica entre viés e discriminação é essencial para que os desenvolvedores e tomadores de decisão compreendam não apenas as nuances técnicas, mas também as implicações sociais mais amplas desses sistemas automatizados. Assim, ressalta-se a importância da elaboração e verificação da representatividade dos bancos de dados e das premissas de decisão dos algoritmos, que devem ser tão rigorosas quanto os possíveis efeitos da discriminação causados pelos modelos.

Autonomia e privacidade

Os algoritmos de personalização coletam dados pessoais dos usuários para criar experiências altamente personalizadas. Esses algoritmos filtram quais informações são apresentadas ao usuário com base na compreensão de preferências, comportamentos e talvez até vulnerabilidades a serem influenciadas – uma linha tênue entre auxiliar e controlar decisões (Mittelstadt *et al.*, 2016). A personalização é nítida no funcionamento das redes sociais. Mas se o que vemos é manipulável, será que temos autonomia para tomar nossas decisões?

Os algoritmos também estão transformando a nossa noção de privacidade (Mittelstadt *et al.*, 2016). Com a crescente utilização de *smartphones*, *smartwatches*, sensores, dispositivos de monitoramento médico, entre outros, manter nossa privacidade tornou-se uma tarefa desafiadora (Karale, 2021). A popularização de dispositivos como

Amazon Echo e *Google Home* levantam questões se esses dispositivos ouvem nossas conversas e quem tem acesso a esses dados (Karale, 2021).

Para Rossetti e Angeluci (2021), a autonomia do indivíduo é desrespeitada quando sua escolha pode estar vinculada a interesses de terceiros. Os usuários podem não estar cientes do grau de detalhamento em que suas informações são coletadas e como são utilizadas pelas empresas. Isso levanta preocupações sobre o consentimento informado e a transparência no uso de dados pessoais. Além disso, a personalização pode criar “bolhas”, já que os usuários são expostos apenas a informações e opiniões que reforçam suas visões existentes. Isso pode levar à polarização e à disseminação de desinformação, uma vez que os algoritmos tentam manter os usuários engajados mostrando-lhes conteúdos que eles já concordam, em vez de oferecer perspectivas diversificadas (Knobloch-Westerwick; Westerwick, 2023).

A onipresença dos algoritmos de personalização desencadeia uma reflexão profunda sobre os limites da nossa capacidade de tomar decisões autônomas. Assim, é imperativo que avancemos em direção a abordagens mais éticas, transparentes e responsáveis na criação e implementação de algoritmos de personalização, assegurando que o progresso tecnológico respeite os princípios fundamentais de autonomia, privacidade e diversidade.

Rastreabilidade e responsabilidade

Tradicionalmente, os desenvolvedores têm controle sobre o funcionamento de algoritmos, na medida em que podem explicar seu design e função a terceiros (Mittelstadt *et al.*, 2016). Esse entendimento de responsabilidade no *design* de *softwares* pressupõe que o programador é capaz de prever possíveis erros e de corrigir resultados indesejáveis (Mittelstadt *et al.*, 2016). No entanto, com o advento dos algoritmos de aprendizado de máquina “caixas pretas”, principalmente dos de aprendizagem não supervisionada, o modelo tradicional de responsabilidade deixa de existir. Não há controle suficiente sobre as tomadas de decisões dos algoritmos, e, portanto, não há quem possa assumir a responsabilidade por elas (Mittelstadt *et al.*, 2016).

Os carros autônomos ilustram bem esta preocupação (Winfield e Jirotko, 2018; Doneda *et al.*, 2018; Stilgoe, 2018). No caso de acidentes envolvendo carros autônomos, quem é o responsável? O(a) programador(a)? A empresa que desenvolveu o carro? Segundo Mittelstadt *et al.* (2016), ainda não há um consenso sobre a responsabilidade dos algoritmos. Porém, há uma crescente opinião de que os algoritmos do futuro precisarão, no mínimo, serem projetados para refletir as normas éticas e culturais de seus usuários e da sociedade (Winfield; Jirotko, 2018).

OS PRINCIPAIS OBSTÁCULOS EM APRENDIZADO DE MÁQUINA

No desenvolvimento de algoritmos de aprendizado de máquina, superficial ou profundo, os(as) programadores(as) precisam evitar algumas armadilhas que podem ter efeitos graves no desempenho dos modelos. Portanto, esta seção abordará cinco obstáculos comuns no desenvolvimento de algoritmos de aprendizado de máquina: dados de baixa qualidade, sobreajuste (*overfitting*), subajuste (*underfitting*), má seleção da métrica de avaliação e o vazamento de dados (*data leakage*).

2.1 DADOS DE BAIXA QUALIDADE

O aprendizado de máquina nos ajuda a encontrar padrões em um conjunto de dados, que são utilizados para fazer previsões sobre novos dados (Google AI, 2023). Como vimos, dados de baixa qualidade são um desafio significativo no desenvolvimento de algoritmos de aprendizado de máquina, uma vez que a qualidade dos dados usados para treinar, validar e testar modelos de aprendizado de máquina tem um impacto direto na precisão, eficácia e confiabilidade dos algoritmos.

Dados de baixa qualidade podem significar: dados insuficientes; dados desbalanceados (existência de categorias sub-representadas); dados incompletos; e *outliers* (pontos fora da curva) (Nex Software Development Company, 2023). Esses dados de baixa qualidade podem levar a resultados imprecisos, viés algorítmico e consequências indesejáveis (modelo pouco acurado ou discriminação). Portanto, é essencial realizar uma análise cuidadosa dos dados antes do desenvolvimento de modelos. De acordo com um curso de preparação de dados para aprendizado de máquina, disponibilizado pelo Google (Google AI, 2023), a construção de um bom conjunto de dados deve se dar conforme o fluxograma da Figura 1.

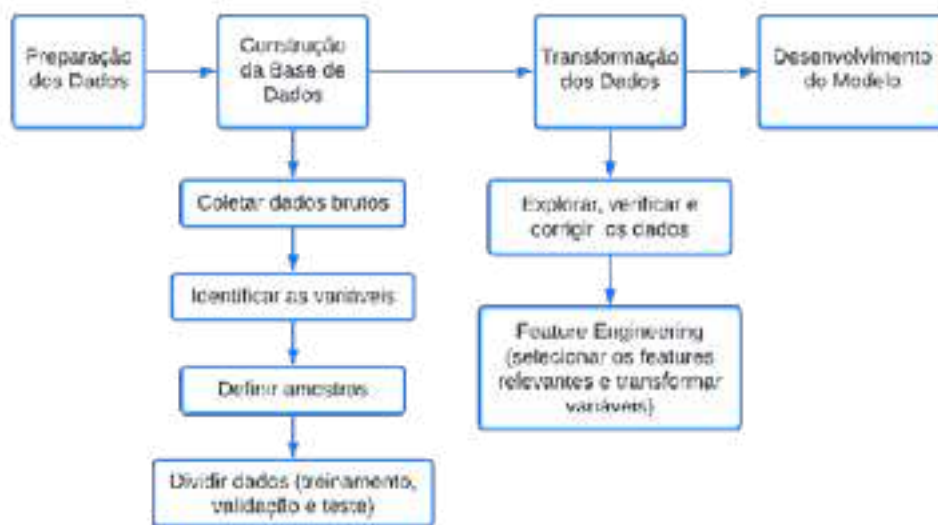


Figura 1 - Fluxograma da preparação de dados

Fonte: Adaptado pelo Autor (Google AI, 2023).

Sobreajuste - *overfitting*

O sobreajuste - *overfitting* - ocorre quando o modelo se ajusta tão bem aos dados de treinamento que perde a capacidade de generalização (Santos *et al.*, 2018). Dessa forma, quando validado junto ao conjunto de teste, o desempenho é significativamente inferior quando comparado ao desempenho com o conjunto de treinamento (Santos *et al.*, 2018). Pode-se dizer, neste caso, que o algoritmo “decorou” os dados de treino, e, portanto, quando os dados de teste são introduzidos, o modelo tenta aplicar as mesmas regras decoradas, obtendo resultados piores (já que os dados são diferentes) (Santos *et al.*, 2018). Segundo Correa (2021), o *overfitting* é comparável a um estudante que estuda para uma prova decorando informações ou exercícios, e quando na prova surgem questões diferentes das decoradas, o estudante não sabe resolvê-las.

A Figura 2 explica a ocorrência do *overfitting*. Na Figura 2, no gráfico à esquerda, os dados estão divididos em dados de treino (azul) e de teste (laranja). A linha pontilhada azul representa a linha de tendência do modelo, considerando os dados de treino. Aparentemente, esta linha também se ajusta aos dados de teste (laranja). Já no gráfico

à direita, a linha de tendência do modelo para os dados de treino (azul), não se ajusta bem aos dados de teste (laranja). Ou seja, o modelo é muito adequado para os dados de treino (azul), mas perde a capacidade de generalização para os dados de teste (laranja), demonstrando o *overfitting*.

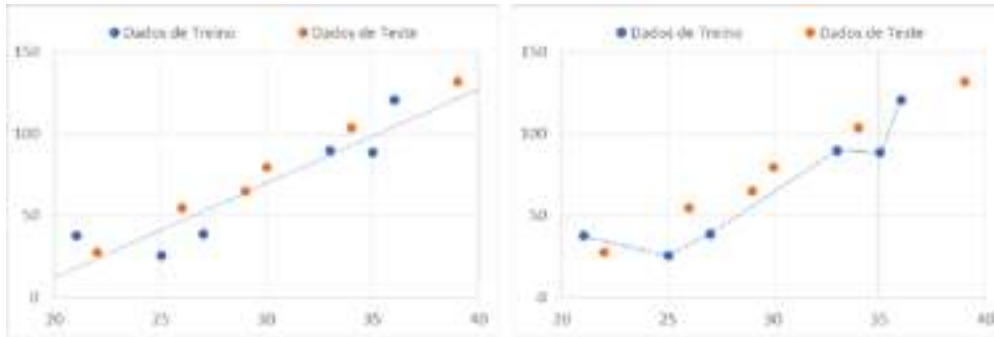


Figura 2 - Sobreajuste (*overfitting*)

Fonte: Elaborado pelos Autores (2023).

Uma das maneiras mais adotadas de se combater o *overfitting* é por meio da validação cruzada (*k-fold*) (Wagner e Rondinelli, 2016). Neste processo, o(a) cientista de dados divide os dados em partições iguais, usando uma fração dos dados para testar o desempenho do modelo e o restante para treinamento, em um processo repetitivo (Santos *et al.*, 2018). Os dados são particionados em *k* subconjuntos, chamados *folds* (Santos *et al.*, 2018). As partições são então iteradas para que cada partição tenha sido usada como conjunto de teste uma vez e os erros sejam calculados (Santos *et al.*, 2018), conforme ilustra a Figura 3. Isso tem o efeito de simular a precisão com que o modelo lidará com novas observações (Wagner e Rondinelli, 2016). Assim, o uso da validação cruzada pode reduzir os vieses (agrupamentos desbalanceados) que podem ocorrer ao se separar simplesmente o banco de dados em conjunto de treino e conjunto de teste.



Figura 3 - Esquema do processo de validação cruzada (k-fold) para k = 10
Fonte: Elaborado pelos Autores

Subajuste - underfitting

Já o underfitting (subajuste) ocorre quando o modelo não é capaz de aprender os padrões presentes nos dados de treinamento de forma adequada. Isso significa que o modelo não consegue se ajustar bem aos dados, resultando em baixo desempenho tanto nos dados de treinamento quanto nos dados de teste. Nesse problema, os resultados tanto dos dados de treinamento quanto dos dados de teste são ruins (Nex Softsys Software Development Company, 2023). Dessa forma, é mais fácil descobrir se o modelo está sofrendo de underfitting do que overfitting, já que no underfitting as métricas de validação serão ruins, ao passo que métricas boas podem tanto significar um bom desempenho geral do modelo quanto overfitting ou a ocorrência de vazamento de dados (que será visto nas próximas seções) (Nex Softsys Software Development Company, 2023).

Uma das formas de combater o underfitting é aumentar o tamanho do conjunto de dados de treino, ou adicionar mais atributos (features) ao modelo, tornando o modelo um pouco mais complexo (Nex Softsys Software Development Company, 2023). A Figura 4 representa uma situação de underfitting. A linha de tendência (em azul) não

representa bem os dados de treino e de teste. Ou seja, o modelo não está prevendo adequadamente o comportamento desse fenômeno, o que pode derivar de diversas razões, como a baixa qualidade do banco de dados, uso de algoritmos com mecanismos incompatíveis com o problema em questão, má calibração dos parâmetros de ajuste dos algoritmos (tunning), entre outros.

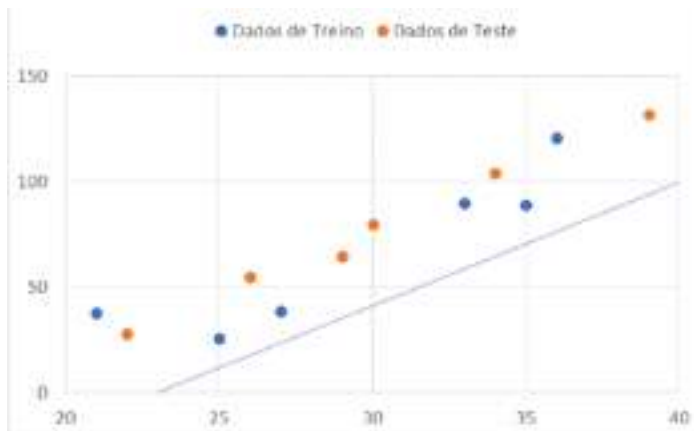


Figura 4 - Linha de tendência subajustada (underfitting)

Fonte: Elaborado pelos Autores

Má seleção da métrica de avaliação

As métricas de avaliação ajudam a avaliar o desempenho dos modelos de aprendizado de máquina. A seleção adequada da métrica de avaliação é uma etapa crucial no desenvolvimento de modelos, pois influencia diretamente as decisões tomadas durante o processo de treinamento e ajuste do modelo. Uma má escolha da métrica de avaliação pode levar a conclusões enganosas sobre o desempenho do modelo e resultar em decisões prejudiciais.

Existem muitos tipos de métricas de avaliação disponíveis (Altexsoft, 2023). No entanto, não existe uma regra a respeito de quais métricas devem ser utilizadas, já que isso depende do tipo de modelo que está sendo desenvolvido (Nex Softsys Software Development Company, 2023). Os modelos de regressão, por exemplo, podem ser avaliados pelas métricas Mse (Erro Quadrado Médio), Rmse (Raiz do Erro Quadrado

Médio), Mae (Erro Médio Absoluto), Mape (Erro médio percentual absoluto), coeficiente de determinação (R^2), entre outras (Altexsoft, 2023). Já os modelos de classificação, que tem como objetivo classificar dados (exemplo: classificar um e-mail como spam ou não-spam), geralmente utilizam métricas como acurácia, precisão, *recall* e *F1-score* (Altexsoft, 2023).

A acurácia, por exemplo, uma das métricas de avaliação mais simples, determina o número de previsões corretas do modelo (Altexsoft, 2023). No entanto, se o conjunto de dados utilizado estiver desbalanceado, alguma categoria estiver sub-representada, essa métrica é inadequada para avaliar o modelo (Nex Softsys Software Development Company, 2023; Altexsoft, 2023). Por exemplo, em um conjunto de treinamento em que 10% dos dados referem-se a pessoas com alguma doença e 90% a pessoas saudáveis, se o modelo não for bem elaborado e classificar todas as pessoas como “saudáveis”, errando a classificação de todos os doentes, a acurácia é de 90% (Altexsoft, 2023). Isso causa uma falsa impressão de bom desempenho do modelo. Portanto, a acurácia não é uma métrica adequada nesse caso (Nex Softsys Software Development Company, 2023).

Vazamento de dados - *data leakage*

O vazamento de dados - *data leakage* - em aprendizado de máquina não está relacionado ao contexto de segurança, mas ao vazamento de informações do conjunto de treinamento para o conjunto de teste (Neoway, 2023). Nesse caso, o desempenho nos conjuntos de teste será semelhante ao obtido nos conjuntos de treinamento, não porque o modelo seja capaz de prever corretamente, mas porque existem padrões semelhantes nas amostras de treinamento e teste (Santos, Soares, *et al.*, 2018). Se o modelo estiver alcançando resultados muito bons em todas as avaliações, pode ser que esteja ocorrendo vazamento de dados (Nex Softsys Software Development Company, 2023). Nesse caso, o modelo é chamado de excessivamente otimista (*overly optimistic*) (Santos, Soares, *et al.*, 2018).

O vazamento de dados pode ser ilustrado com a situação de um aluno que está

estudando para uma prova e tem acesso ao gabarito (Neoway, 2023). O conhecimento do gabarito certamente influenciará os seus estudos. Provavelmente o aluno terá ótimos resultados na prova, mas isso não significa que ele estará preparado para aplicar o conhecimento em situações diferentes (Neoway, 2023). No aprendizado de máquina é semelhante: se ocorre o vazamento de dados, há um bom desempenho de treinamento e teste, mas não há um bom desempenho em situações reais (Nex Softsys Software Development Company, 2023). Portanto, estratégias rigorosas devem ser impreterivelmente adotadas na divisão dos dados em dados de treinamento e teste.

As melhores práticas em ia

O desenvolvimento acelerado da Inteligência Artificial nos últimos anos e o seu potencial impacto em diversos setores da sociedade estão provocando um amplo debate sobre os princípios e valores que devem guiar o seu desenvolvimento (Jobin, Ienca e Vayena, 2019). O conhecimento das implicações legais, éticas e de segurança do uso da IA são fundamentais para grupos de pesquisa e entidades públicas e privadas na atualidade (Koshiyama *et al.*, 2021; Google AI, 2022). Para Koshiyama *et al.* (2021), assim como na auditoria financeira, eventualmente governos, empresas e sociedade exigirão a auditoria de algoritmos, ou seja, a garantia formal de que os algoritmos são legais, éticos e seguros.

No que diz respeito à pesquisa científica, não se tem conhecimento, até o momento, de uma diretriz universal específica para pesquisa científica em IA publicada por comitês de ética que se aplique a todos os contextos e jurisdições. No entanto, as considerações éticas em torno da pesquisa em IA estão se tornando cada vez mais importantes. Nesse sentido, várias instituições e órgãos reguladores publicaram relatórios e documentos a respeito das melhores práticas a serem adotadas no desenvolvimento das IAs nos últimos anos (Jobin, Ienca e Vayena, 2019). Organizações profissionais, como a Association for Computing Machinery (ACM) e o Institute of Electrical and Electronics Engineers (IEEE), desenvolveram diretrizes éticas e códigos de conduta

para pesquisadores de IA (Mittelstadt, 2019). Essas diretrizes geralmente enfatizam princípios como transparência, responsabilidade, justiça e prevenção de danos (Mittelstadt, 2019). No entanto, segundo Mittelstadt (2019), esses documentos permanecem comparativamente curtos, teóricos e carentes de aconselhamento e normas comportamentais específicas.

Em 2019, a Comissão Europeia nomeou um Grupo de Especialistas de Alto Nível em Inteligência Artificial para elaboração do documento “Orientações Éticas para uma IA de Confiança” (High-Level Expert Group On AI, 2019). De acordo com o documento, uma IA de confiança deve ser: legal - garantindo o respeito à legislação existente, ética - observando princípios e valores éticos; e sólida do ponto de vista técnico e social. O documento também prescreve quatro princípios éticos imprescindíveis no desenvolvimento das IAs: respeito pela autonomia humana, prevenção de danos, justiça e explicabilidade. A partir desses princípios, o documento elabora sete requisitos que devem ser aplicados neste contexto, conforme ilustra a Figura 5.



Figura 5 - Requisitos para uma IA de confiança

Fonte: Adaptado de High-Level Expert Group On AI, 2019.

O Google (2023) também disponibiliza em seu site “Google AI” uma lista de recomendações a serem adotadas no desenvolvimento de algoritmos de IA. Nesta lista há recomendações gerais, para o desenvolvimento de algoritmos justos, para garantir a

interpretabilidade, para garantir a privacidade e para garantir a segurança. Essas recomendações serão abordadas a seguir de forma resumida.

Recomendações gerais

- usar uma abordagem de design centrado no ser humano: a experiência de usuários(as) reais com o sistema é essencial para avaliar o verdadeiro impacto de suas previsões, recomendações e decisões. Recomenda-se a utilização de recursos de design que garantam uma boa experiência do(a) usuário(a) e a incorporação de feedbacks antes e durante o desenvolvimento do projeto;

- utilizar várias métricas para avaliar o modelo: o uso de várias métricas ajudará a entender diferentes tipos de erros e experiências;

- examinar diretamente os dados brutos: os modelos refletirão os dados em que são treinados, portanto, sempre que possível, deve-se analisar os dados brutos para verificar a existência de dados faltosos, dados desbalanceados, dados redundantes, possíveis vieses, entre outros - sempre respeitando-se a privacidade;

- entender as limitações do conjunto de dados e do modelo: um modelo treinado para detectar correlações estatísticas não deve ser utilizado para fazer inferências causais. Embora duas variáveis possam estar correlacionadas entre si, isso não significa que uma variável seja a causa da outra. Portanto, é importante entender o escopo e as limitações dos modelos, bem como comunicar as limitações aos usuários;

- testar o modelo várias vezes: testar os modelos de aprendizado de máquina superficial e profunda é uma prática crucial para garantir a eficácia, robustez e segurança dos algoritmos em aplicações reais. Portanto, recomenda-se a realização de testes rigorosos para testar cada componente do sistema isoladamente e testes de integração para entender como os componentes do sistema interagem com outras partes;

- monitorar e atualizar os modelos: atualizar e monitorar o desempenho dos modelos de aprendizado de máquina são práticas fundamentais para garantir que os algoritmos de IA continuem eficazes e relevantes ao longo do tempo.

Recomendações para um algoritmo justo

- projetar o modelo para que seja imparcial: o envolvimento de uma equipe multidisciplinar no desenvolvimento de algoritmos, como advogados(as), cientistas sociais, engenheiros(as), cientistas da computação, entre outros, é essencial para que se leve em conta diferentes perspectivas;
- usar conjunto de dados representativos: recomenda-se o uso de dados representativos para treinar e testar os modelos ou deixar claro ao(a) usuário(a) e demais desenvolvedores quando existirem grupos sub-representados;
- verificar a existência de vieses: recomenda-se selecionar conjuntos de dados de teste diversificados que possam testar o sistema de forma adversa. Além disso, recomenda-se atualização contínua dos sistemas e o encorajamento de feedback de usuários;
- analisar o desempenho: recomenda-se avaliar o desempenho continuamente e em um amplo espectro de usuários.

Recomendações para garantir a interpretabilidade

- planejar as ações para evitar a opacidade: trabalhar em colaboração com diferentes especialistas contribui para identificação de quais recursos de interpretabilidade são necessários e por quê;
- tratar a interpretabilidade como uma parte central da experiência do usuário: recomenda-se a interação com os usuários no ciclo de desenvolvimento para testar e refinar suposições sobre as necessidades e metas do usuário;
- projetar o modelo para ser interpretável: recomenda-se a utilização do menor número de entradas possíveis para que se entenda os fatores que influenciam o modelo e o uso preferencial de algoritmos que permitam a interpretabilidade dos cálculos, decisões e etapas intermediárias de desenvolvimento;

- comunicação com o usuário: é importante que os modelos ofereçam explicações compreensíveis e adequadas para os diferentes usuários – sempre que possível.

Recomendações para garantir a privacidade

- coletar e manipular os dados com responsabilidade: recomenda-se que os modelos sejam treinados sem o uso de dados confidenciais. Se não for possível, os dados confidenciais devem ser manuseados com cuidado e deve-se respeitar a legislação existente;

- proteger a privacidade dos dados: recomenda-se verificar se o modelo está memorizando ou expondo involuntariamente dados confidenciais ou permitindo inferências e associações.

Recomendações para garantir a segurança

- identificar possíveis ameaças ao sistema: recomenda-se o desenvolvimento de um modelo de ameaças rigoroso para entender todos os vetores de ataque possíveis;

- desenvolver uma abordagem para combater ameaças: testar o desempenho dos sistemas em cenários adversos e ter um plano em caso de problemas;

- continuar aprendendo: manter-se atualizado sobre os últimos avanços na área é imprescindível para melhorias no desempenho de defesas.

Há algumas críticas a respeito desses documentos que propõe diretrizes para o desenvolvimento de IAs éticas. Segundo Larsson (2020), o documento “Orientações Éticas para uma IA de Confiança” da Comissão Europeia enfatiza questões de responsabilidade, transparência e proteção de dados no desenvolvimento de tecnologias de IA confiáveis, mas não aborda a interação entre a legislação e essas diretrizes. De acordo

com Jobin, Ienca e Vayene (2019), traduzir os princípios éticos para a prática e aliar a ética à legislação são próximos passos importantes para a comunidade global.

Jobin, Ienca e Vayene (2019) analisaram 84 documentos que tratavam sobre princípios e diretrizes para o desenvolvimento de tecnologias de IA éticas e observaram que há uma convergência global em torno de cinco princípios éticos: transparência, justiça e equidade, não-maleficência, responsabilidade e privacidade. No entanto, existem divergências relacionadas à interpretação, importância e implementação desses princípios. Essas diferenças podem prejudicar a criação de uma agenda global para IA ética, já que há incerteza sobre quais princípios priorizar e como resolver conflitos entre eles (Jobin, Ienca e Vayena, 2019). Por exemplo, a necessidade de conjuntos de dados cada vez maiores e diversificados para evitar vieses pode entrar em conflito com a exigência de dar aos indivíduos maior controle sobre seus dados, a fim de respeitar sua privacidade e autonomia. Além disso, os autores observaram que os países do norte global, principalmente os EUA, têm uma representação proeminente no debate a respeito da IA ética. Essa sub-representação do sul-global levanta preocupações a respeito da equidade e justiça global na discussão da ética em IA (Jobin, Ienca e Vayena, 2019).

O debate da ética em IA deve envolver uma força de trabalho diversificada, capaz de incorporar conhecimentos críticos variados e de verificar possíveis vieses (GOOGLE AI, 2023). Segundo o Grupo de Especialistas de Alto Nível em Inteligência Artificial, que elaborou o documento “Orientações Éticas para uma IA de Confiança”, mais do que a definição de um conjunto de regras, assegurar a confiança na IA requer a construção e manutenção de uma cultura e uma mentalidade ética por meio do envolvimento em debates públicos, da educação e da aprendizagem prática (High-Level Expert Group On AI, 2019).

Considerações finais

Com o rápido e crescente desenvolvimento da IA nas mais variadas áreas de estudo, entender seu impacto social e implicações éticas se tornou uma necessidade da sociedade atual. Nesse sentido, este artigo abordou as principais questões éticas relacionadas à IA, os principais obstáculos no desenvolvimento de algoritmos de aprendizado

de máquina superficial e profundo, e as melhores práticas para desenvolver algoritmos éticos, interpretáveis, seguros e justos.

Vieses e preconceitos podem se perpetuar facilmente através do uso de dados desbalanceados e de correlações infundadas. Diante disso, a colaboração entre desenvolvedores de algoritmos e outros especialistas (especialmente cientistas sociais) torna-se fundamental para compreender diversas perspectivas e identificar as sutis formas de propagação dos preconceitos. Além disso, é imprescindível o uso de conjuntos de dados representativos e balanceados para o treinamento e teste dos modelos de aprendizado de máquina.

A opacidade dos algoritmos é um dos principais obstáculos para a aplicação e aceitação das técnicas de IA. A falta de transparência e auditabilidade em algoritmos que tomam decisões governamentais, mediam transações bancárias ou fornecem diagnósticos médicos parece injusta e preocupante. A compreensão lógica dessas decisões é essencial para detectar falhas e atribuir responsabilidades. Portanto, é urgente a implementação de governança ética, transparente, inclusiva e ágil pelas organizações que desenvolvem e operam as IAs, já que a ampla aceitação futura desses algoritmos dependerá dessas questões. Assim, é possível que, em breve, exija-se que os algoritmos sejam auditáveis, a fim de garantir a ética, legalidade e segurança dos modelos. Pesquisadores(as) e desenvolvedores(as) devem se preparar para esse cenário o quanto antes.

A ética dos algoritmos não é uma questão que será solucionada apenas através de uma abordagem tecnológica. Essa temática envolve também assuntos de ordem social, cultural, jurídica e política. A implantação de algoritmos éticos requer uma abordagem holística, que leve em conta não apenas a eficácia e a precisão dos modelos, mas também suas implicações sociais e impactos na sociedade como um todo. O desenvolvimento tecnológico e a responsabilidade social devem caminhar lado a lado, a fim de evitar que as diferenças sociais cresçam ainda mais. Portanto, mais do que discutir os princípios éticos da IA, é preciso que esses princípios sejam colocados em prática por meio das legislações.

Por fim, com o rápido desenvolvimento da Inteligência Artificial e dos algoritmos de aprendizado de máquina, algumas das técnicas apresentadas nesse artigo podem se tornar obsoletas. Por isso, recomenda-se a contínua atualização a respeito das melhores práticas e avanços científicos relacionados ao tema.

Agradecimentos

agradecemos à CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Código 001 – bolsa de doutorado para Carolina Lopes) e à FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais, projeto número APQ-01838-21 para Julia Mendes) pelo apoio financeiro. Agradecemos também a colaboração do Grupo de Pesquisa em Ciência de Dados aplicada à Engenharia (CIDENG – CNPq).

Referências

ADAMSON, Adewole S.; SMITH, Avery. *Machine Learning and Health Care Disparities in Dermatology*. *JAMA Dermatology*, v. 154, n. 11, p. 1247-1248, 2018.

ALTEXSOFT. *Machine Learning Metrics: How to Measure the Performance of a Machine Learning Model*, 16 jun. 2023. Disponível em: <https://www.altexsoft.com/blog/machine-learning-metrics/>. Acesso em: 26 abr. 2023.

CORREA, Danielle M. *Feature Engineering: preparando dados para aprendizado de máquina*. Ateliware, 12 maio 2021. Disponível em: <https://ateliware.com/blog/feature-engineering>. Acesso em: 29 jun. 2023.

GARCIA, Ana C. *Ética e Inteligência Artificial*. *Computação Brasil*, v. 43, p. 14-22, 2020.

GOOGLE AI. *Data Preparation and Feature Engineering in ML*, 2023. Disponível em: <https://developers.google.com/machine-learning/data-prep>. Acesso em: 11 maio 2023.

GOOGLE AI. *Responsible AI practices*, 2023. Disponível em: <https://ai.google/responsibilities/responsible-ai-practices/>. Acesso em: 17 maio 2023.

HIGH-LEVEL EXPERT GROUP ON AI. *Ethics guidelines for trustworthy AI*.

JOBIN, Anna; IENCA, Marcello; VAYENA, Effy. *The global landscape of AI ethics guidelines*. *Nature Machine Intelligence*, v. 1, p. 389-399, 2019.

KARALE, Ashwin. *The challenges of IoT addressing security, ethics, privacy, and laws. Internet of Things*, v. 15, p. 100420, 2021.

KNOBLOCH-WESTERWICK, Silvia; WESTERWICK, Axel. *Algorithmic personalization of source cues in the filter bubble: Self-esteem and self-construal impact information exposure. New Media & Society*, v. 25, n. 8, 2023.

KOSHIYAMA, Adriano et al. *Towards algorithm auditing: a survey on managing legal, ethical and technological risks of AI, ML and associated algorithms. SSRN*, 2021.

LARSSON, Stefan. *On the Governance of Artificial Intelligence through Ethics Guidelines. Asian Journal of Law and Society*, v. 7, n. 3, p. 437-451, 2020.

MITTELSTADT, Brent. *Principles alone cannot guarantee ethical AI. Nature Machine Intelligence*, v. 1, p. 501-507, 2019.

MITTELSTADT, Brent D. et al. *The ethics of algorithms: Mapping the debate. Big Data & Society*, v. 3, n. 2, p. 1-21, 2016.

NEOWAY. *O que é data leakage e como essa falha pode contaminar as decisões*, 12 ago. 2023. Disponível em: <https://blog.neoway.com.br/data-leakage/>. Acesso em: 04 jul. 2023.

NEX SOFTSYS SOFTWARE DEVELOPMENT COMPANY. *Common Pitfalls Which May Mislead Results In Machine Learning*, 2023. Disponível em: <https://www.nexsoftsys.com/articles/common-pitfalls-in-machine-learning.html>. Acesso em: 13 abr. 2023.

OBERMEYER, Ziad. et al. *Dissecting racial bias in an algorithm used to manage the health of populations. Science*, v. 366, n. 6464, p. 447-453, 2019.

ROSSETTI, Regina; ANGELUCI, Alan. *Ética Algorítmica: questões e desafios éticos do avanço tecnológico da sociedade da informação. Galáxia*, v. 46, p. 1-18, 2021.

SANTOS, Miriam S. et al. *Cross-validation for imbalanced datasets: avoiding overoptimistic and overfitting approaches. IEEE Computational Intelligence Magazine*, v. 13, n. 4, p. 59-76, 2018.

STILGOE, Jack. *Machine learning, social learning and the governance of self-driving cars*. *Social Studies of Science*, v. 48, n. 1, p. 25-26, 2018.

WAGNER, Nicholas; RONDINELLI, James M. *Theory-guided machine learning in materials science*. *Frontiers in Materials*, v. 3, p. 28, 2016.

WINFIELD, Alan F.; JIROTKA, Marina. *Ethical governance is essential to building trust in robotics and artificial intelligence systems*. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, v. 376, n. 2133, p. 20180085, 2018.



■ Giselle Beiguelman, frames do vídeo generativo da série Flora Rebellis, versão 2024.
Imagens geradas com Inteligência Artificial (Style Gans)

■ [Acesse o vídeo clicando na imagem](#)





INTELIGÊNCIA ARTIFICIAL, BLOCKCHAIN E A CADEIA DE CUSTÓDIA DA PROVA NO PROCESSO PENAL

DOI: <https://doi.org/10.35699/2965-6931.2023.47605>

MATHEUS ARAÚJO*

RESUMO: O presente estudo visa discutir o emprego da Inteligência Artificial e Blockchain no procedimento da cadeia de custódia de provas e vestígios digitais. A pesquisa é de natureza qualitativa e a metodologia utilizada consistiu em método hipotético-dedutivo, por meio da revisão bibliográfica de autores que versam sobre inteligência artificial, tecnologia, cadeia de custódia da prova e direito processual penal. Assim, será feita uma análise do funcionamento da Inteligência Artificial e do Blockchain, bem como serão examinados os principais aspectos da cadeia de custódia, a fim de constatar a compatibilidade desta com as referidas tecnologias. Concluiu-se, ao final, pela possibilidade de utilização da Inteligência Artificial e do Blockchain para auxiliar no registro e preservação da cadeia de custódia, no âmbito do direito brasileiro, diante da segurança que esse tipo de tecnologia fornece.

PALAVRAS-CHAVE: inteligência artificial; cadeia de custódia; provas digitais; blockchain; direito processual penal.

Artificial intelligence, blockchain and the chain of custody in criminal procedure

ABSTRACT: This study aims to discuss the use of Artificial Intelligence (AI) and Blockchain in the Chain of Custody of Digital Evidence and traces. The research is qualitative, and the methodology is based on a hypothetical-deductive method, through the bibliographical review of authors who deal with AI, Blockchain, technology, and criminal procedural law. Thus, an analysis of the functioning of artificial intelligence and blockchain will be carried out, as well as the main aspects of the chain of custody will be examined, in order to verify its compatibility with the referred technologies. It was concluded, in the end, for the possibility of using AI and Blockchain to assist in the registration and preservation of the chain of custody, within the scope of Brazilian law, given the security that this type of technology provides.

KEYWORDS: artificial intelligence; chain of custody; digital evidence; blockchain; criminal procedure law.

* Universidade Federal de Minas Gerais

1 A definição legal de Cadeia de Custódia está prevista no art. 158-A do Código de Processo Penal, qual seja, “[...] o conjunto de todos os procedimentos utilizados para manter e documentar a história cronológica do vestígio coletado em locais ou em vítimas de crimes, para rastrear sua posse e manuseio a partir de seu reconhecimento até o descarte.” (BRASIL, 2019).

2 Cumpre aqui fazer uma breve diferenciação entre vestígios e provas, uma vez que sua distinção é extremamente relevante. O art. 158-A, § 3º, do Código de Processo Penal define vestígio como “todo objeto ou material bruto, visível ou latente, constatado ou recolhido, que se relaciona à infração penal.” Prova, por sua vez, são meios utilizados para a reconstrução da história de um fato. Lopes Jr. (2019, p. 425) destaca as principais características da prova, quais sejam, “1. estão dirigidos a convencer o juiz de uma afirmação; 2. estão a serviço do processo e integram o processo penal; 3. dirigem-se a formar a convicção do juiz para o julgamento final - tutela de segurança; 4. servem à sentença; 5. exigem estrita observância da publicidade, contradição e imediação; 6. são praticados ante o juiz que julgará o processo.”

Introdução

O presente artigo tem como objetivo discutir o emprego da Inteligência Artificial (IA) e do *Blockchain* na preservação de provas digitais, bem como na Cadeia de Custódia da Prova no Processo Penal Brasileiro. Para tanto, o objeto deste trabalho será limitado para responder às seguintes perguntas: (i) o ordenamento jurídico brasileiro autoriza o uso da Inteligência Artificial e do *Blockchain* na preservação de provas digitais?; (ii) como a Inteligência Artificial pode ser empregada em harmonia com os princípios e regras que regem o processo penal brasileiro?; e (iii) como a Inteligência Artificial e o *Blockchain* podem ser aplicadas ao procedimento da cadeia de custódia? Ao final, pretende-se demonstrar que o uso da inteligência artificial e do *Blockchain* é um importante meio para aprimorar o instituto da cadeia de custódia da prova, principalmente quando tratar-se da preservação de vestígios e provas digitais.

No tocante à metodologia de pesquisa empregada, a abordagem do problema parte de uma pesquisa de natureza qualitativa, sendo adotado o método de abordagem indutivo, que será desenvolvido a partir da pesquisa bibliográfica de livros e artigos científicos que versam sobre os seguintes conceitos: Processo Penal; Cadeia de Custódia; Vestígios Digitais; Metadados; Inteligência Artificial; *Machine Learning*, *Blockchain* e Cadeia de Custódia.

A hipótese adotada nesta pesquisa é a de que a Inteligência Artificial e o *Blockchain* podem ser empregados no procedimento da cadeia de custódia, especialmente quando se tratar de vestígios e provas digitais, diante da possibilidade de análise dos metadados pela inteligência artificial, com o escopo de identificar eventuais alterações nos vestígios digitais e, posteriormente, o *Blockchain* serviria para preservar a integridade e a autenticidade das informações contidas nas provas digitais, que são mais suscetíveis de manipulação e, conseqüentemente, demandam proteção maior do que os vestígios físicos. Para fundamentar a referida hipótese pretende-se utilizar as obras de Geraldo Prado, Janaína Matida e Aury Lopes Jr., no intuito de traçar as principais características

do instituto da cadeia de custódia, bem como os princípios aplicáveis ao instituto, a fim de verificar sua compatibilidade com as tecnologias mencionadas. Também serão consultadas as obras de Leonardo Parentoni, Marina Chatterjee, Vishal Maini e Karl Wüst, com o objetivo de compreender o funcionamento da inteligência artificial e do *blockchain*, de modo a constatar as formas de aplicação dessas tecnologias no âmbito da cadeia de custódia das provas digitais. Outras fontes e autores serão consultados ao longo desta pesquisa.

Ao final, pretende-se demonstrar a compatibilidade da IA e do *Blockchain* com a cadeia de custódia da prova digital no ordenamento jurídico brasileiro, demonstrando sua aplicação no procedimento de preservação de vestígios digitais, descrevendo, ainda, as vantagens que essas tecnologias oferecem na manutenção da integridade e autenticidade das provas digitais. Por outro lado, serão enumerados os cuidados que devem ser observados no emprego desse tipo de tecnologia, a fim de não comprometer princípios que regem o processo penal brasileiro.

³ PRADO, Geraldo. A cadeia de custódia da prova no processo penal. 2. ed. Rio de Janeiro: Marcial Pons, 2021, p. 144.

A cadeia de custódia da prova no processo penal brasileiro

A Cadeia de Custódia¹ foi positivada no ordenamento jurídico brasileiro por meio da Lei nº 13.964/19, que introduziu os artigos 158-A ao 158-F no Código de Processo Penal. Os referidos dispositivos legais descrevem uma série de procedimentos que devem ser adotados pelos órgãos encarregados pela persecução penal, a fim de se preservar os vestígios² e provas encontradas ao longo de uma investigação. Prado define a cadeia de custódia como um procedimento necessário para “saber se um determinado elemento probatório está em condições de ser avaliado, ou seja, se o elemento probatório pode ser objeto de avaliação (...)”³. O objetivo, ao final, é assegurar a integridade e a autenticidade dos elementos colhidos durante a fase de investigação até o final do processo. A integridade pode ser entendida como algo que se encontra inteira, isto é, que não sofreu qualquer diminuição ou perda de sua inteireza. A autenticidade, por sua vez, corresponde ao caráter de legitimidade do vestígio, que não teve seu conteúdo adulterado ou corrompido.

4 LOPES JR., Aury. *Direito processual penal*. 17. ed. São Paulo: Saraiva Educação, 2020, p. 411.

5 LOPES JR., Aury. *Direito processual penal*. 17. ed. São Paulo: Saraiva Educação, 2020, p. 412.

6 “O Estado não deve ser superior ao criminoso apenas pelo ângulo do monopólio do exercício legítimo da coação. Sua superioridade antes de qualquer coisa deve ser ética.” PRADO, Geraldo. *A cadeia de custódia da prova no processo penal*. 2. ed. Rio de Janeiro: Marcial Pons, 2021, p. 216.

7 PRADO, Geraldo. *A cadeia de custódia da prova no processo penal*. 2. ed. Rio de Janeiro: Marcial Pons, 2021, p. 194.

8 Não é o foco desse trabalho discutir as consequências jurídicas da quebra da cadeia de custódia, por isso optou-se por uma abordagem breve deste tema. Recomenda-se a leitura dos autores mencionados nas notas de rodapé para se aprofundar acerca do tema.

9 O entendimento mencionado é sustentado por Gustavo Badaró e Del-tan Dallagnol. Badaró explica que: “Por outro lado, no caso de vícios mais graves, em que se tenham dúvidas sobre a autenticidade ou a integridade da fonte de prova, em que haja uma probabilidade de que ela tenha sido adulterada, substituída ou modificada, isso enfraquecerá seu valor, cabendo ao julgador, motiva-

A doutrina identifica dois princípios que regem a cadeia de custódia, quais sejam a mesmidade e a desconfiança. Segundo Lopes Jr., o primeiro princípio estabelece que “a prova valorada é exatamente e integralmente aquela que foi colhida, correspondendo, portanto, “a mesma”⁴, ou seja, o procedimento visa assegurar a integridade do vestígio, desde o momento de sua coleta até a sua apresentação em juízo, de modo a evitar alterações no vestígio ao longo do percurso. O princípio da desconfiança, por sua vez, “consiste na exigência de que a prova (documentos, DNA, áudios etc.) deva ser “acreditada”, submetida a um procedimento que demonstre que tais objetos correspondem ao que a parte alegar ser.”⁵ O referido princípio decorre da lógica de que não se deve atribuir, automaticamente, confiança aos elementos de prova colhidos ao longo de uma investigação pelo simples fato de ser o Estado⁶ atuando, muito pelo contrário, compete à autoridade policial e ao Ministério Público comprovarem que a não houve perda ou adulteração de um vestígio. Portanto, a integridade e a autenticidade de uma prova não são presumidas, mas sim comprovadas por meio da cadeia de custódia, de modo que “durante a investigação é que deve ficar demonstrado que o elemento probatório não foi manipulado pelo seu detentor original e pelos agentes do Estado responsáveis pela custódia, e que, portanto, está preservado e íntegro à disposição da defesa e do juiz.”⁷

A consequência jurídica⁸ da violação da cadeia de custódia é bastante discutida na doutrina, havendo, pelo menos, três entendimentos acerca de sua inobservância. O primeiro entendimento⁹ é de que a violação da cadeia de custódia, por si só, não acarreta a inadmissibilidade do elemento de prova, mas, tão somente, o enfraquecimento de seu valor probatório no momento de sua avaliação pelo juízo. De outro lado, o segundo entendimento¹⁰ defende que a violação da cadeia de custódia implica na inadmissibilidade do elemento de prova, de modo que este não poderá ser avaliado pelo juízo, devendo ser excluído do processo. Por fim, identifica-se um entendimento intermediário¹¹, que avalia o grau de violação da cadeia de custódia para estabelecer se haverá a inadmissibilidade da prova ou o enfraquecimento de seu valor no momento do julgamento. A posição defendida por Prado e Matida parece a mais adequada, especialmente quando se tratar de vestígios digitais, que são mais suscetíveis de sofrer alterações e modificações no seu conteúdo.

O procedimento da cadeia de custódia

Conforme mencionado anteriormente, a cadeia de custódia é um procedimento descrito nos artigos 158-A ao 158-F do Código de Processo Penal, sendo que os referidos dispositivos legais não esgotam o tema, tampouco estabelecem quais as técnicas que devem ser empregadas para preservação dos elementos de prova encontrados ao longo da investigação, o que abre margem para adoção de diversas técnicas, especialmente o emprego de meios tecnológicos para se atingir os fins pretendidos pelo legislador.

Em síntese, a cadeia de custódia¹² tem início com a identificação de vestígios que sejam de interesse para a apuração de um crime, sendo que os vestígios deverão ser descritos conforme se encontram no local, sendo possível o registro em fotografia e filmagem. Identificados os vestígios, proceder-se-á ao isolamento do ambiente, a fim de evitar a alteração do estado em que se encontra. Após o isolamento do local, os vestígios serão coletados e acondicionados em embalagens individualizadas e lacradas, contendo anotação da data, hora e o nome do responsável pela coleta e acondicionamento do vestígio. Dessa forma, os vestígios devidamente acondicionados serão transportados até a central de custódia, onde serão recebidos. No momento do recebimento dos vestígios, é de suma importância que haja a documentação do número de procedimento, da autoridade policial responsável, local de origem, nome de quem transportou o vestígio, código de rastreamento, o tipo de vestígio e identificação de quem o recebeu. Após o recebimento, será possível romper o lacre da embalagem que condiciona o vestígio, a fim de realizar o seu processamento, isto é, os exames periciais e a manipulação dos vestígios para a extração das informações relevantes para a investigação. Por fim, o vestígio será armazenado, em condições adequadas, até que haja autorização judicial para o seu descarte.

Os dispositivos legais que regulam a cadeia de custódia atentam para a importância de se registrar todas as pessoas que tiveram contato com o vestígio ao longo da cadeia de custódia, bem como o motivo do contato, o que foi feito com o vestígio na etapa de processamento e os métodos empregados para a preservação do vestígio. Matida explica que “os novos dispositivos assentam que a confiabilidade dependerá do necessário oferecimento de um detalhado rastreio, nos moldes definidos normativamente”.¹³ A documentação¹⁴ de todo o procedimento é essencial para que, posteriormente, as partes possam fazer a análise da cadeia de custódia, a fim de constatar eventuais desvios

damente, fazer tal análise. Evidente que, em tal caso, se se considerar atendível um meio de prova, decorrente de fonte sobre a qual haja irregularidades na cadeia de custódia, haverá uma inegável necessidade de reforço justificativo demonstrando o porquê ser possível confiar na autenticidade e integridade de tal fonte.” BADARÓ, Gustavo Henrique. *Processo penal*. 8. ed. rev., atual. e ampl. São Paulo: Thomson Reuters Brasil, 2020, p. 514-515.

10 Essa corrente é defendida por Geraldo Prado, Janaína Matida e Aury Lopes Jr.

11 Guilherme Madeira Dezem esclarece que “a consequência da violação da cadeia de custódia pode ser a nulidade da prova, sua ilicitude ou simplesmente o enfraquecimento da força probante deste meio de prova. A consequência dependerá da violação havida na cadeia de custódia.” DEZEM, Guilherme Madeira. *Curso de processo penal*. 8. ed. rev., atual. e ampl. São Paulo: Thomson Reuters Brasil, 2021, p. 733.

12 As fases da cadeia de custódia estão descritas no art. 158-B do Código de Processo Penal: Art. 158-B. A cadeia de custódia compreende o rastreamento do vestígio nas seguintes etapas:

I - reconhecimento: ato de distinguir um elemento como de potencial interesse para a produção da prova pericial;

II - isolamento: ato de evitar que se altere o estado das coisas, devendo isolar e preservar o ambiente imediato, mediato e relacionado aos vestígios e local de crime;

III - fixação: descrição detalhada do vestígio conforme se encontra no local de crime ou no corpo de delito, e a sua posição na área de exames, podendo ser ilustrada por fotografias, filmagens ou croqui, sendo indispensável a sua descrição no laudo pericial produzido pelo perito responsável pelo atendimento;

IV - coleta: ato de recolher o vestígio que será submetido à análise pericial, respeitando suas características e natureza;

V - acondicionamento: procedimento por meio do qual cada vestígio coletado é embalado de forma individualizada, de acordo com suas características físicas, químicas e biológicas, para posterior análise, com anotação da data, hora e nome de quem realizou a coleta e o acondicionamento;

VI - transporte: ato de transferir o vestígio de um local para o outro, utilizando as condições adequadas (embalagens, veículos, temperatura, entre outras), de modo a garantir a manutenção de suas características originais, bem como o controle de sua posse;

VII - recebimento: ato formal de transferência da posse do vestígio, que deve ser documentado com, no mínimo, informações referentes ao número de procedimento e unidade de polícia judiciária relacionada, local de origem, nome de quem transportou o vestígio, código de rastreamento

ou violações ao longo do procedimento. A ausência de um relatório impede a análise judicial após o término das investigações, o que representa um prejuízo para as partes, principalmente para a defesa do investigado, que sequer tem o conhecimento da trajetória percorrida pelo elemento de prova.

A cadeia de custódia nas provas digitais

Uma breve leitura dos dispositivos legais que regulam a cadeia de custódia no Código de Processo Penal permitiria inferir que o procedimento somente se aplicaria aos chamados vestígios materiais, isto é, as evidências físicas encontradas na investigação. Contudo, a referida interpretação não se mostra adequada, de modo que é perfeitamente possível a aplicação da cadeia de custódia nas chamadas provas digitais¹⁵. Inclusive, mostra-se necessária a adoção do referido procedimento, tendo em vista a fragilidade¹⁶ que esse tipo de vestígio apresenta em relação aos vestígios materiais, diante da maior facilidade de exclusão e adulteração das informações contidas em meios digitais.

Há um consenso na doutrina acerca da possibilidade de aplicação da cadeia de custódia nos vestígios e provas digitais, bem como os tribunais superiores¹⁷ já reconhecem a necessidade de adoção do procedimento da cadeia de custódia nas provas digitais. Não restam dúvidas, portanto, da importância de se observar o procedimento dos artigos 158-A ao 158-F para assegurar a confiabilidade de uma prova digital, sob pena de inadmissibilidade no processo.

Acerca dos métodos e técnicas utilizados na preservação de provas digitais, cumpre demonstrar o tratamento do tema nos Estados Unidos da América, uma vez que existe, no âmbito federal, um guia¹⁸ para auxiliar as autoridades policiais nas investigações que encontrarem vestígios digitais. O referido guia contém uma série de deveres que devem ser seguidos pelo primeiro policial que entrar em contato com a cena do crime, especialmente quando se deparar com evidências digitais. Destaca-se, assim, algumas das providências e cuidados¹⁹ previstos no guia: assegurar, imediatamente, todos os dispositivos eletrônicos encontrados na cena; impedir que pessoas não autorizadas tenham contato com os dispositivos eletrônicos; remover todas as pessoas não autorizadas da cena do crime; garantir que a condição dos dispositivos eletrônicos não seja

alterada; deixar os dispositivos eletrônicos da forma em que foram encontrados, isto é, ligados ou desligados; não desligar dispositivos eletrônicos que estejam ligados; verificar se existe alguma comunicação, transmissão ou envio de informações nos dispositivos ligados, bem como se há algum processo em andamento para excluir conteúdo do dispositivo etc.

Além disso, há uma forte preocupação em documentar²⁰ a cena do crime para criar um registro para a investigação, de modo que se recomenda a filmagem e fotografia do local para uma melhor compreensão da cena do crime, bem como a posição e o local em que cada um dos dispositivos eletrônicos foram encontrados. Há, inclusive, recomendações de procurar por certos tipos de evidências digitais a depender do delito investigado, o que demonstra uma atenção especial dada a esse tipo de prova por parte das autoridades policiais norte americanas.

Verifica-se, portanto, que os vestígios e provas digitais demandam um maior cuidado na cadeia de custódia, pois, conforme dito anteriormente, são mais suscetíveis a manipulações e adulterações, tendo em vista a maior volatilidade que as informações digitais apresentam. Certamente, todo o cuidado é pouco quando o assunto é preservar vestígios digitais, de modo que toda e qualquer técnica voltada para esta finalidade é bem-vinda e deve ser incentivada no âmbito de qualquer investigação.

A (in)compatibilidade da aplicação da inteligência artificial e do blockchain no procedimento da cadeia de custódia de vestígios digitais

Feita uma explicação do que é cadeia de custódia e os seus objetivos, passa-se a responder, no presente tópico, os dois primeiros problemas apresentados na introdução deste trabalho. Em síntese, pretende-se demonstrar a compatibilidade da inteligência artificial e do *Blockchain* com o procedimento da cadeia de custódia previsto no ordenamento jurídico brasileiro.

É comum que tecnologias inovadoras sejam vistas com certa resistência por parte dos indivíduos, especialmente no âmbito do direito, onde parecer haver uma resistên-

to, natureza do exame, tipo do vestígio, protocolo, assinatura e identificação de quem o recebeu;

VIII - processamento: exame pericial em si, manipulação do vestígio de acordo com a metodologia adequada às suas características biológicas, físicas e químicas, a fim de se obter o resultado desejado, que deverá ser formalizado em laudo produzido por perito;

IX - armazenamento: procedimento referente à guarda, em condições adequadas, do material a ser processado, guardado para realização de contraperícia, descartado ou transportado, com vinculação ao número do laudo correspondente;

X - descarte: procedimento referente à liberação do vestígio, respeitando a legislação vigente e, quando pertinente, mediante autorização judicial.

13 MATIDA, J. A cadeia de custódia é condição necessária para a redução dos riscos de condenações de inocentes. Revista da Defensoria Pública do Estado do Rio Grande do Sul, Porto Alegre, n. 27, p. 17–26, 2021, p. 21. Disponível em: <https://revista.defensoria.rs.def.br/defensoria/article/view/269>. Acesso em: 26 jun. 2023.

14 Prado descreve o que deve estar contido na documentação da cadeia de custódia: “a) os protocolos de coleta/extração do dado ou elemento probatório, de sorte a comprovar que não houve supressão, inclusão ou alteração de elementos

que afete a qualidade de prova autêntica e íntegra, em atenção ao princípio da mesmidade; b) os cuidados que foram adotados, voltados à transparência do processo anterior, que igualmente demonstrem que houve controle por terceiros da coleta/extração, acondicionamento, transporte e preservação do elemento probatório, isso em reverência ao princípio da desconfiança; c) a cadeia de pessoas que tiveram contato com o elemento probatório e os respectivos títulos e motivos, se funcionário público ou terceiro à administração, se para o transporte ou exame do vestígio etc.” PRADO, Geraldo. A cadeia de custódia da prova no processo penal. 2. ed. Rio de Janeiro: Marcial Pons, 2021, p. 168-169.

¹⁵ Vaz conceitua prova digital como “os dados em forma digital (no sistema binário) constantes de um suporte eletrônico ou transmitidos em rede de comunicação, os quais contêm representação de fatos ou ideias.” VAZ, Denise Provasi. Provas digitais no processo penal: formulação do conceito, definição das características e sistematização do procedimento probatório. 2012. Tese de Doutorado. Universidade de São Paulo, p. 64.

¹⁶ Prado explica que: “Na atualidade, toda a literatura sobre provas digitais alerta para os graves riscos de manipulação ou perda da integridade dos elementos probatórios, com prejuízo a sua autenticidade

maior com o emprego de tecnologias nas práticas forenses. Contudo, não se pode renunciar às vantagens oferecidas pelos avanços tecnológicos ao direito nas últimas décadas, desde que em sintonia com os princípios e regras vigentes no ordenamento jurídico brasileiro.

Antes de adentrar nas razões pelas quais a IA e o *Blockchain* são compatíveis com a cadeia de custódia, destaca-se, brevemente, os conceitos das referidas tecnologias. Segundo Maini e Sabri²¹, a Inteligência Artificial pode ser conceituada como “o estudo de agentes que percebem o mundo ao seu redor, formam planos e tomam decisões para alcançar seus objetivos.” Neste sentido, Chatterjee²² destaca que “a inteligência artificial visa permitir que as máquinas executassem o raciocínio, replicando a inteligência humana. Como o principal objetivo dos processos de IA é ensinar as máquinas a partir da experiência, alimentar as informações corretas e a autocorreção é crucial.” Trata-se, portanto, de uma tecnologia cujo objetivo é reproduzir a inteligência humana, a partir do reconhecimento e análise de informações, de modo a possibilitar o desenvolvimento de algoritmos que permitam a tomada autônoma de decisões e execução de ações. Tem-se como exemplo o ChatGPT, que consiste em “um modelo de linguagem baseado na arquitetura GPT-3.5, desenvolvido pela OpenAI, que utiliza redes neurais e técnicas avançadas de processamento de linguagem natural para gerar respostas coerentes e contextualmente relevantes em diálogos por texto.”²³ Para fornecer as respostas nos diálogos, o ChatGPT é alimentado com dados textuais obtidos pela internet, reconhecendo padrões e obtendo informações relevantes para proporcionar a resposta mais adequada ao diálogo iniciado pelo usuário do *software*.

O *Blockchain*, por sua vez, “é uma base de dados descentralizada, que é mantida por uma rede distribuída de computadores”.²⁴ Acerca de seu funcionamento, Wüst e Gervais explicam que “cada bloco está ligado ao bloco anterior por meio de um hash²⁵ criptográfico. Um bloco é uma estrutura de dados que permite armazenar uma lista de transações. As transações são criadas e trocadas por pares da rede *Blockchain* e modificam o estado da *Blockchain*”²⁶. O melhor exemplo para compreender a tecnologia *Blockchain* são as criptomoedas, que dependem da referida tecnologia para efetuar transações envolvendo ativos digitais. “Para que uma transação de *bitcoins* seja efetuada, o protocolo determina que todas as transações anteriores sejam verificadas, a começar

pelo primeiro bloco, o ‘*genesis block*’”.¹ Dessa forma, cada bloco armazena as mesmas informações relativas as transações de *bitcoins*, validando as operações para confirmar a transferência.

A simples descrição dos conceitos acima permite identificar diversas utilidades que as tecnologias supramencionadas podem oferecer à cadeia de custódia. Assim, diante das possibilidades oferecidas pela inteligência artificial e pelo *Blockchain*, haveria alguma vedação por parte do ordenamento jurídico brasileiro no emprego das referidas tecnologias na cadeia de custódia? A resposta é negativa, explica-se.

O processo penal brasileiro é regido pelos princípios da ampla defesa e do contraditório², que estabelecem, em síntese, que as partes poderão utilizar de todos os meios e recursos lícitos para se defenderem, bem como terão o direito de conhecer e se manifestar acerca de todos os atos processuais praticados, a fim de influenciar na decisão final a ser tomada pelo julgador. Além disso, conforme já mencionado, a Cadeia de Custódia é regida pelos princípios da mesmidade e desconfiança, que estabelecem um ônus para os órgãos investigadores de demonstrarem, por meio de um relatório, que os elementos de prova colhidos ao longo de uma investigação foram devidamente

de, e isso não é relevante pelo ordenamento jurídico ao considerar as consequências que advêm da constatação da imprecisão técnica e das condições de preservação de elementos probatórios desta natureza.” PRADO, Geraldo. A cadeia de custódia da prova no processo penal. 2. ed. Rio de Janeiro: Marcial Pons, 2021, p. 192.

17 Destaca-se trecho do voto do Ministro Ribeiro Dantas, da Quinta Turma do Superior Tribunal de Justiça, no julgamento do Agravo Regimental em Habeas Corpus nº 143.169/RJ: “Quando entram em cena as fontes de prova imateriais, ou aquelas que, conquanto tenham um suporte físico, são essencialmente intangíveis (a exemplo dos dados informáticos), não é diferente: em observância às peculiaridades dessas espécies probatórias, há técnicas específicas que precisam ser adotadas pelo aparato sancionador para garantir objetivamente a confiabilidade das provas por ele produzidas.” BRASIL. Superior Tribunal de Justiça. Quinta Turma. Agravo Regimental em Habeas Corpus nº 143.169/RJ. Agravante: R.L.S.M. Agravado: Ministério Público do Estado do Rio de Janeiro. Advogado: Marcelo Turbay Freiria. Rel. Min. Jesuino Rissato. Brasília, 07 de fev. de 2023. Disponível em: <https://ww2.stj.jus.br/processo/pesquisa/>. Acesso em: 26 jun. 2023.

I MOREIRA, Athur Salles de Paula; DELGADO, Camila Campos Baumgratz; SANTOS, Gabriel Gonçalves. Repensando a tecnologia *Blockchain*: por que nem tudo o que você leu até hoje era verdade? In: PARENTONI, Leonardo; MILAGRES, Marcelo de Oliveira; VAN DE GRAAF, Jeroen (Coords.). MOREIRA, Arthur Salles de Paula; CHAGAS, Ciro Costa; SANTANA, Mariana Damiani (Orgs). *Direito, Tecnologia e Inovação*, v. III: *Aplicações Jurídicas de Blockchain*. Expert Editora: Belo Horizonte, 2021, p. 29.

2 “A parte deve ter conhecimento sobre os atos processuais praticados no processo e deve poder reagir a esses atos, impugnando-os”. DEZEM, Guilherme Madeira. Curso de processo penal. 8. ed. rev., atual. e ampl. São Paulo: Thomson Reuters Brasil, 2021, p. 132-133. Ainda sobre o contraditório, Badaró esclarece que: “A incidência da igualdade substancial no contraditório mostrou a necessidade de implementá-lo e efetivá-lo. O contraditório deixa de ser mera possibilidade para se transformar em uma realidade. Deve haver real e igualitária participação dos sujeitos processuais ao longo de todo o processo, assegurando a efetividade e a plenitude do contraditório. É o que se denominou contraditório efetivo e equilibrado.” BADARÓ, Gustavo Henrique. Processo penal. 8. ed. rev., atual. e ampl. São Paulo: Thomson Reuters Brasil, 2020, p. 61.

18 U.S. DEPARTMENT OF JUSTICE OFFICE OF JUSTICE. *Electronic Crime Scene Investigation: A Guide for First Responders*. Second Edition. Programs National Institute of Justice. Disponível em: <https://www.ncjrs.gov/pdffiles1/nij/219941.pdf>. Acesso em: 26 jun. 2023.

19 U.S. DEPARTMENT OF JUSTICE OFFICE OF JUSTICE. *Electronic Crime Scene Investigation: A Guide for First Responders*. Second Edition. Programs National Institute of Justice, p. 15-16. Disponível em: <https://www.ncjrs.gov/pdffiles1/nij/219941.pdf>. Acesso em: 26 jun. 2023.

20 “Documentation of a crime scene creates a record for the investigation. It is important to accurately record the location of the scene; the scene itself; the state, power status, and condition of computers, storage media, wireless network devices, mobile phones, smart phones, PDAs, and other data storage devices; Internet and network access; and other electronic devices. The first responder should be aware that not all digital evidence may be in close proximity to the computer or other devices.” U.S. DEPARTMENT OF JUSTICE OFFICE OF JUSTICE. *Electronic Crime Scene Investigation: A Guide for First Responders*. Second Edition. Programs National Institute of Justice, p. 19. Disponível em: <https://www.ncjrs.gov/pdffiles1/nij/219941.pdf>. Acesso

preservados sem nenhuma perda de conteúdo, tampouco adulteração de suas informações. Há, portanto, uma desconfiança na atuação do estado investigador, cabendo a este comprovar a idoneidades dos vestígios colhidos no curso da investigação. “A preservação das fontes de prova é concebida como remédio jurídico-processual contra o desequilíbrio inquisitório, caracterizado pela seleção e uso arbitrário de elementos pelas agências repressivas”³. Consta-se que o registro da cadeia de custódia é essencial para a efetivação dos princípios da ampla defesa e contraditório, bem como para assegurar o devido processo legal⁴, uma vez que possibilita ao acusado ter pleno conhecimento de toda a história cronológica da prova submetida à apreciação das partes e do julgador.

O princípio da desconfiança reforça a ideia de que os órgãos investigadores devem adotar todos os meios possíveis para fazer um registro mais fidedigno de todo o percurso traçado pelo vestígio até a sua apresentação em juízo. Sendo assim, todo e qualquer método que aprimorar a preservação dos vestígios e o registro de todo o procedimento deverá ser empregado, especialmente quando se constata que os dispositivos legais que versam sobre o tema não estabelecem nenhuma técnica particular que deve ser utilizada no procedimento, tampouco vedam o uso de alguma tecnologia específica. Infere-se, portanto, que o legislador optou por permitir o uso de qualquer método, desde que seja voltado para aprimorar as finalidades de preservar os vestígios e documentar sua história cronológica.

Estabelecidas às premissas que regem o processo penal e a cadeia de custódia, constata-se que a Inteligência Artificial e o *Blockchain* poderiam ser utilizados para promover uma maior efetividade aos princípios mencionados anteriormente. Conforme será demonstrado no próximo capítulo, a IA pode ser utilizada para analisar metada-

3 PRADO, Geraldo. *A cadeia de custódia da prova no processo penal*. 2. ed. Rio de Janeiro: Marcial Pons, 2021, p. 191.

4 “Em suma, o modelo constitucional do devido processo legal no sistema brasileiro é de um processo que se desenvolva perante o juiz natural, em contraditório, assegurada a ampla defesa, com atos públicos e decisões motivadas, em que ao acusado seja assegurada a presunção de inocência, devendo o processo se desenvolver em um prazo razoável. Sem isso, não haverá *due process* ou um processo equo.” BADARÓ, Gustavo Henrique. *Processo penal*. 8. ed. rev., atual. e ampl. São Paulo: Thomson Reuters Brasil, 2020, p. 97.

em: 26 jun. 2023.

dos associados a arquivos digitais, por exemplo, informações referentes à criação ou exclusão de mensagens de texto ou áudios, o que permitiria verificar a autenticidade dos arquivos, detectando possíveis adulterações de conteúdo. Já o *Blockchain*, por possuir uma natureza imutável, possibilitaria o registro de informações no relatório da Cadeia de Custódia, sem que os registros pudessem ser alterados, fornecendo, assim, um histórico confiável e verificável acerca de todo o percurso realizado pelo vestígio digital até o seu descarte. Todas essas vantagens oferecidas por essas tecnologias fortalecem os princípios da mesmidade e desconfiância, bem como colaboram para o exercício da ampla defesa e do contraditório, pois fornecem uma maior transparência e lisura ao procedimento, possibilitando que as partes tenham um conhecimento mais preciso de todo o caminho traçado pelo vestígio e, a partir deste conhecimento, é possível a realização de um contraditório efetivo.

A plicações práticas da inteligência artificial e do blockchain na cadeia de custódia de provas digitais

Demonstrada a possibilidade de utilização da inteligência artificial e do *Blockchain* na cadeia de custódia, passa-se a descrever algumas aplicações práticas das referidas tecnologias no procedimento da cadeia de custódia. Neste momento, serão abordadas algumas funcionalidades que a inteligência artificial e o *Blockchain* podem oferecer no objetivo de preservar a integridade e a autenticidade de vestígios e provas digitais.

A inteligência artificial aplicada à cadeia de custódia

Já a IA, como dito anteriormente, visa replicar a inteligência humana, por meio do reconhecimento e análise de informações para a tomada de decisões autônomas. Esse

21 MAINI, Vishal; SABRI, Samer. Machine Learning for humans. 2017, p. 9. Disponível em: <https://everythingcomputerscience.com/books/Machine%20Learning%20for%20Humans.pdf>. Acesso em: 28 abr. 2023. No original: "Artificial intelligence is the study of agents that perceive the world around them, form plans, and make decisions to achieve their goals." (MAINI; SABRI, 2017, p. 9).

22 CHATTERJEE, Marina. Data Science vs Machine Learning and Artificial Intelligence: the difference explained, 2023. Disponível em: <https://www.mygreatlearning.com/blog/difference-data-science-machine-learning-ai/>. Acesso em: 28 abr. 2023. No original: "artificial intelligence aims at enabling machines to execute reasoning by replicating human intelligence. Since the main objective of AI processes is to teach machines from experience, feeding the correct information and self-correction is crucial."

23 OpenAI. OpenAI's GPT-3: Language Models are Unsupervised Multitask Learners. Disponível em: <https://openai.com/research/gpt-3/>. Acesso em: 28 abr. 2023.

24 MOREIRA, Athur Salles de Paula; DELGADO, Camila Campos Baumgratz; SANTOS, Gabriel Gonçalves. Repensando a tecnologia Blockchain: por que nem tudo o que você leu até hoje era verdade? In: PARENTONI, Leonardo; MILAGRES, Marcelo de Oliveira; VAN DE GRAAF, Jeroen (Coords.). MOREIRA, Athur Salles de Paula; CHAGAS, Ciro Costa; SANTANA, Mariana Damiani (Orgs). Direito, Tecnologia e Inovação, v. III: Aplicações Jurídicas de Blockchain. Expert Editora: Belo Horizonte, 2021, p. 26.

25 Destaca-se o conceito de hash fornecido por Schneier: "One-way hash functions are like digital fingerprints: small pieces of data that can serve to identify much larger digital objects. They are public functions; no secret keys are involved." SCHNEIER, Bruce. Secret & Lies: Digital Security in a Networked World. Indianapolis:

26 WÜST, Karl; GERVAIS, Arthur. Do you need a Blockchain? 2018 crypto valley conference on blockchain technology (CV-CBT). IEEE, 2018, p. 46. No original: "The name Blockchain stems from its technical structure - a chain of blocks. Each block is linked to the previous block with a cryptographic hash. A block is a data-structure which allows to store a list of transactions. Transactions are created and exchanged by peers of the Blockchain network

aprendizado pode se dar de várias maneiras, sendo denominado de *Machine learning* ou Aprendizado de Máquina. O Aprendizado de Máquina nada mais é do que um subconjunto da inteligência artificial, que possibilita a esta aprender por conta própria. Dessa forma, "um algoritmo de *Machine Learning* permite identificar padrões em dados analisados, construir modelos que expliquem o mundo e prever acontecimentos sem ter que estar previamente programada por regras e modelos"³¹. Um exemplo deste uso são os aplicativos de navegação, no qual utilizam a referida técnica de aprendizado para traçar as rotas mais rápidas para se chegar ao endereço pretendido pelo usuário. Verifica-se que não há programação prévia, uma vez que o aplicativo, por conta própria, a partir da análise de dados, direciona o usuário para o caminho mais rápido.

Existem diversos tipos de aprendizados de máquina³², contudo, não é o objetivo deste trabalho abordar de forma exaustiva todas as modalidades existentes, de modo que será tratado apenas as formas de aprendizagem que, no entendimento do autor, se mostram mais adequadas para os fins pretendidos pela cadeia de custódia.

Destaca-se, inicialmente, o aprendizado supervisionado, que é um tipo de algoritmo no qual um modelo é treinado a partir de um conjunto de dados rotulados, de modo que cada exemplo é composto por uma entrada e sua saída correspondente. O objetivo é encontrar uma função que mapeie as entradas para as saídas corretas, permitindo ao modelo fazer previsões precisas em dados não vistos.³³ Em síntese, no aprendizado supervisionado, a inteligência artificial é exposta a diversos exemplos previamente rotulados, com entradas e saídas, e, a partir dessas informações, o modelo se adapta para fazer previsões acerca de novos exemplos que forem surgindo no curso de seu funcionamento. Cita-se como exemplo a classificação de e-mails entre "spam" e "não spam". Neste caso, o conjunto de treinamento consistiria em vários e-mails, onde cada e-mail é representado com algumas características (por exemplo, palavras-chave, comprimento do e-mail etc.) e possui um rótulo indicando se é um spam ou não. O modelo é treinado para aprender a distinguir entre os dois tipos de e-mails com base nas características fornecidas. O aprendizado supervisionado apresenta duas tarefas, quais sejam, regressão e classificação. A primeira tem como objetivo prever um valor numérico contínuo com base em um conjunto de variáveis de entrada. A classificação, por sua vez, tem a função de "atribuir um rótulo"³⁴, isto é, a partir de das características de um determinado objeto, a inteligência artificial classifica este objeto, fazendo as devidas distinções.

No âmbito da cadeia de custódia, o aprendizado supervisionado poderia ser utilizado para identificar adulterações em provas digitais, em uma fase incipiente do procedimento, no intuito de averiguar eventuais manipulações do vestígio digital, antes deste ser submetido a perícia. Dessa forma, um modelo de aprendizado supervisionado pode ser treinado usando um conjunto de dados rotulados que contenham exemplos de vestígios digitais autênticos e exemplos com adulterações conhecidas. Têm-se como exemplos de adulteração as alterações em metadados³⁵, edições em arquivos de imagem ou manipulações em registros de atividades. A exposição da inteligência artificial aos dados rotulados como adulterados permitiria que o modelo desenvolvesse um algoritmo capaz de identificar modificações indevidas em vestígios digitais.

O uso da Inteligência Artificial na fase incipiente do procedimento da cadeia de custódia tem como objetivo identificar e eliminar, desde o princípio, vestígios digitais que possam ter sofrido algum tipo de adulteração por terceiros, com o propósito de incriminar ou isentar alguém de responsabilidade. Assim, seria possível identificar eventuais manipulações indevidas ocorridas na fase de reconhecimento e coleta do dispositivo eletrônico. Apenas para exemplificar, imagina-se que, no curso de uma investigação, os policiais se deparam com um celular que pode conter informações relevantes para o crime em apuração. Dessa forma, o celular é coletado, acondicionado em uma embalagem devidamente lacrada e transportado até a central de custódia. Na central de custódia, após o rompimento do lacre, antes de ser iniciada a perícia no dispositivo telefônico, este seria submetido a IA, no intuito de identificar alguma adulteração do seu conteúdo. Caso a IA constate alguma adulteração no vestígio digital, tal fato deverá ser registrado na Cadeia de Custódia e o referido vestígio deverá ser descartado, não podendo compor o acervo de evidências utilizado para embasar a investigação, tampouco poderá ser utilizado como prova na fase processual.

O aprendizado aprofundado³⁶, por sua vez, poderia ser utilizado para proteger a privacidade das pessoas investigadas. Esse tipo de aprendizado utiliza de redes neurais artificiais para aprender e representar dados de forma hierárquica, por meio de camadas consecutivas de neurônios artificiais, que processam e transformam os dados de entrada, repassando as informações relevantes para as camadas seguintes. A vantagem desse tipo de técnica é possibilitar o processamento de imagens, vídeos, áudios, reconhecimento de falas, detecção de objetos etc.³⁷ Como exemplo de aprendi-

and modify the state of the Blockchain. As such, transactions can exchange monetary amounts, but are not restricted to financial transactions only and for example allow to execute arbitrary code within so-called smart contracts.”

27 MOREIRA, Athur Salles de Paula; DELGADO, Camila Campos Baumgratz; SANTOS, Gabriel Gonçalves. Repensando a tecnologia Blockchain: por que nem tudo o que você leu até hoje era verdade? In: PARENTONI, Leonardo; MILAGRES, Marcelo de Oliveira; VAN DE GRAAF, Jeroen (Coords.). MOREIRA, Arthur Salles de Paula; CHAGAS, Ciro Costa; SANTANA, Mariana Damiani (Orgs.). Direito, Tecnologia e Inovação, v. III: Aplicações Jurídicas de Blockchain. Expert Editora: Belo Horizonte, 2021, p. 29.

28 “A parte deve ter conhecimento sobre os atos processuais praticados no processo e deve poder reagir a esses atos, impugnando-os”. DEZEM, Guilherme Madeira. Curso de processo penal. 8. ed. rev., atual. e ampl. São Paulo: Thomson Reuters Brasil, 2021, p. 132-133. Ainda sobre o contraditório, Badaró esclarece que: “A incidência da igualdade substancial no contraditório mostrou a necessidade de implementá-lo e efetivá-lo. O contraditório deixa de ser mera possibilidade para se transformar em uma realidade. Deve haver real e igualitária participação dos sujeitos proces-

suais ao longo de todo o processo, assegurando a efetividade e a plenitude do contraditório. É o que se denominou contraditório efetivo e equilibrado.” BADARÓ, Gustavo Henrique. Processo penal. 8. ed. rev., atual. e ampl. São Paulo: Thomson Reuters Brasil, 2020, p. 61.

29 PRADO, Geraldo. A cadeia de custódia da prova no processo penal. 2. ed. Rio de Janeiro: Marcial Pons, 2021, p. 191.

30 “Em suma, o modelo constitucional do devido processo legal no sistema brasileiro é de um processo que se desenvolva perante o juiz natural, em contraditório, assegurada a ampla defesa, com atos públicos e decisões motivadas, em que ao acusado seja assegurada a presunção de inocência, devendo o processo se desenvolver em um prazo razoável. Sem isso, não haverá due process ou um processo equo.” BADARÓ, Gustavo Henrique. Processo penal. 8. ed. rev., atual. e ampl. São Paulo: Thomson Reuters Brasil, 2020, p. 97.

31 MAINI, Vishal. SABRI, Samer. Machine Learning for humans, 2017, p. 9. Disponível em: <https://everythingcomputerscience.com/books/Machine%20Learning%20for%20Humans.pdf>. Acesso em: 28 abr. 2023. No original: “A machine’s learning algorithm enables it to identify patterns in observed data, build models that explain the world, and predict things without having explicit pre-program-

zado aprofundado, pode-se citar plataformas de recomendação de conteúdo, que utilizam o aprendizado aprofundado para analisar o comportamento do usuário, bem como suas preferências, com o objetivo de apresentar recomendações personalizadas. Mencionam-se, ainda, veículos autônomos e assistentes virtuais como tecnologias que empregam o aprendizado aprofundado para tomar decisões e exercer atividades de forma autônoma. O aprendizado aprofundado permite o desenvolvimento de um algoritmo capaz de analisar e extrair dados de vestígios digitais, identificando informações que estejam relacionadas com o objeto da investigação criminal, evitando uma devassa desnecessária no conteúdo que não guarda relação com crime investigado.

Haveria, portanto, um filtro das informações contidas em um dispositivo eletrônico, de modo que a inteligência artificial, por meio do aprendizado aprofundado, seria capaz de identificar quais das informações são de fato relevantes para a investigação, excluindo-se da análise os dados irrelevantes. Por exemplo, seria possível identificar em um celular quais fotos estão relacionadas ao fato delituoso e quais não, evitando, assim, que os investigadores tenham contato desnecessário com fotos que não têm serventia para investigação, mas que podem de alguma forma, violar a privacidade³⁸ do indivíduo.

Verifica-se, assim, duas contribuições relevantes que a inteligência artificial pode fornecer para a cadeia de custódia, especialmente na fase inicial do procedimento, sendo que as referidas funcionalidades fortalecem os princípios da mesmidade e desconfiança, bem como auxiliam na proteção da privacidade de pessoas investigadas.

Blockchain aplicado a cadeia de custódia

Conforme mencionado anteriormente, o *Blockchain*³⁹ funciona “como uma espécie de ‘livro razão’ no qual todas as informações e eventuais transações são registradas”.

⁴⁰ Vale dizer que o *Blockchain* não se restringe às transações financeiras, podendo ser utilizado para o registro de outras informações. Verifica-se, assim, sua importante contribuição para a Cadeia de Custódia, especialmente no que diz respeito ao dever de registro⁴¹ da cadeia de custódia imposto aos órgãos encarregados pela persecução penal.

Wüst e Gervais⁴² enumeram as principais características do *Blockchain*, quais sejam, a verificabilidade pública; transparência; privacidade; integridade; redundância; e âncora de confiança. Dentre as características mencionadas anteriormente, destacam-se a verificabilidade pública, a transparência e a integridade, uma vez que remetem aos princípios da mesmidade e desconfiança.

A verificabilidade pública e a transparência representam a possibilidade de qualquer pessoa verificar as transações registradas no *Blockchain*, tendo em vista que essas são visíveis para todos os participantes da rede, permitindo, assim, que as informações sejam auditadas publicamente. Além disso, o fato de todos os integrantes da rede terem uma cópia completa das informações registradas facilita a detecção de eventuais disparidades.

A integridade, por sua vez, impede que as informações registradas na rede sofram alterações não autorizadas, garantindo a imutabilidade⁴³ do conteúdo. Alterar ou excluir uma transação registrada requer um consenso da maioria dos nós da rede, o que demandaria um esforço hercúleo para a modificação das informações previamente registradas, sendo praticamente impossível a realização de tal tarefa.

Essas características contidas no *Blockchain* vão ao encontro dos fundamentos da Cadeia de Custódia, uma vez que corroboram com o objetivo de preservação da integridade e autenticidade da prova, por meio do registro de todo o percurso realizado pelo vestígio até a sua apresentação em juízo. Tais funcionalidades permitem um relatório mais fidedigno do caminho traçado pela prova, pois asseguram um registro imutável de todos os acontecimentos da cadeia, sem o risco de alterações que possam ocultar desvios eventualmente ocorridos no percurso, além de possibilitar que as partes, especialmente o acusado, tenham um conhecimento exato de todo o sucedido.

O *Blockchain* pode ser utilizado para o registro dos vestígios digitais obtidos no curso de uma investigação, de modo que cada vestígio será associado a um identificador único, protegido por criptografia. Além do registro dos vestígios, é possível, ainda, o registro de todos os protocolos relativos à coleta das evidências, bem como os cuidados tomados para a preservação dos vestígios, por exemplo, a forma de acondicionamento e transporte até a central de custódia. Outro aspecto essencial que deve ser registrado diz respeito aos indivíduos que tiveram contato com os vestígios, devendo constar a qualificação e o motivo do contato, bem como todos os procedimentos aos quais o ves-

med rules and models.” (MAINI, SABRI, 2017, p. 9).

32 Dentre as diversas técnicas de aprendizado de máquina destacam-se o aprendizado supervisionado, aprendizado não-supervisionado, aprendizado por reforço e o aprendizado profundo. Para um maior aprofundamento sobre o tema, recomenda-se a leitura do artigo Machine Learning for humans, bem como as demais obras citadas neste trabalho acerca da inteligência artificial.

33 KOTSIANTIS, So-tiris B. et al. Supervised machine learning: A review of classification techniques. Emerging artificial intelligence applications in computer engineering, v. 160, n. 1, p. 3-24, 2007.

34 MAINI, Vishal. SABRI, Samer. Machine Learning for humans, 2017, p. 9. Disponível em: <https://everythingcomputerscience.com/books/Machine%20Learning%20for%20Humans.pdf>. Acesso em: 28 abr. 2023. No original: “Classification: Assign a label. Is this a picture of a cat or a dog?” (MAINI, SABRI, 2017, p. 18).

35 Segundo Takahashi, metadados são “dados a respeito de outros dados, ou seja, qualquer dado usado para auxiliar na identificação, descrição e localização de informações. Trata-se em outras palavras, de dados estruturados que descrevem as características de um recurso de informação.”

TAKAHASHI, Tadao. Sociedade da Informação no Brasil: o livro verde. Brasília: Ministério da Ciência e Tecnologia, 2000, p. 173.

36 CHATTERJEE, Marina. Deep Learning Tutorial: What it Means and what's the role of Deep Learning, 2023. Disponível em: <https://www.mygreatlearning.com/blog/what-is-deep-learning/#one>. Acesso em: 17 jun. 2023. Comenta que: "The initial levels help the machines learn simple information, and as the levels increase, the information keeps building. With each new level machines pick up further information and combines it with what it had learnt in the last level. At the end of the process, the system gathers a final piece of information which is a compound input. This information passes through several hierarchies and has semblance to complex logical thinking."

37 LECUN, Yann; BENGIO, Yoshua; HINTON, Geoffrey. Deep learning. Nature, v. 521, n. 7553, p. 437, 2015.

38 PRADO, Geraldo. A cadeia de custódia da prova no processo penal. 2. ed. Rio de Janeiro: Marcial Pons, 2021, p. 184. Comenta que: Em vez de realizar-se uma cópia eletrônica de todo o dispositivo de armazenamento, é preferível, de acordo com as circunstâncias, que os inspetores façam o uso de técnicas forenses que, por exemplo, lhes permitam restringir o universo de pesquisa e apreensão ape-

tigio foi submetido. A título de exemplificação, no caso de um dispositivo eletrônico que será submetido a um procedimento de extração de dados, tal procedimento deverá ser registrado no *Blockchain*, descrevendo todos os métodos empregados, os resultados obtidos e, ainda, a qualificação dos responsáveis seriam escrituradas e rastreadas, criando um histórico transparente de todas as interações havidas com o vestígio.

No que tange o controle de acesso, o *Blockchain*, por meio de chaves criptográficas e permissões, pode controlar o acesso aos vestígios digitais, de modo que somente as partes autorizadas poderiam ter contato com os vestígios digitais registrados na rede. Trata-se de excelente forma de evitar contatos desnecessários de pessoas não autorizadas, aumentando a confiabilidade⁴⁴ do vestígio.

Constata-se que o emprego do *Blockchain* aumenta a transparência e a confiança no procedimento de coleta e manipulação dos vestígios digitais, permitindo que as partes envolvidas possam verificar a autenticidade das provas e ter confiança de que as informações não foram adulteradas ou corrompidas. Há, portanto, um fortalecimento na confiabilidade e no valor probatório dos vestígios digitais, que poderão ser admitidos como prova no âmbito do processo penal pela realização da extração. É de suma importância que sejam anotados na rede todas as pessoas que entraram em contato com o vestígio, tendo em vista que as principais adulterações ocorrem quando a evidência é manipulada por alguém. Daí a necessidade de sempre registrar todos os indivíduos que participaram do procedimento, a fim de permitir a identificação de eventual responsável pela adulteração do vestígio. Dessa forma, todas as ações realizadas no *Blockchain*

Considerações finais

O presente trabalho foi elaborado com o objetivo de solucionar três problemas. Dessa forma, buscou-se responder se ordenamento jurídico brasileiro admitiria o uso da inteligência artificial e do *Blockchain* no procedimento da cadeia de custódia de provas e vestígios digitais; de que forma as referidas tecnologias poderiam se compatibilizar com os princípios e regras que regem o processo penal; e, por fim, descrever al-

gumas aplicações práticas da Inteligência Artificial e *Blockchain* na cadeia de custódia.

Inicialmente, foi feita uma análise do procedimento da cadeia de custódia previsto no Código de Processo Penal brasileiro. O objetivo principal dessa parte do trabalho foi compreender o que é a cadeia de custódia, os seus objetivos e os princípios que regem o referido instituto processual. Restou demonstrado que se trata de procedimento que visa assegurar a integridade e autenticidade dos vestígios obtidos no curso de uma investigação, por meio do registro de todas as etapas percorridas pelo vestígio até a sua apresentação no processo, como prova. Constatou-se a importância da cadeia de custódia nas provas e vestígios digitais, uma vez que esse tipo de evidência é mais suscetível de sofrer adulterações no seu conteúdo.

Posteriormente, passou-se a responder as duas primeiras indagações feitas na introdução deste trabalho. Nessa parte, foi feita uma breve análise dos conceitos de IA e Blockchain, a fim de compreender as referidas tecnologias. Na sequência, passou-se a demonstrar a compatibilidade das referidas tecnologias com os princípios da ampla defesa e do contraditório, bem como os princípios da mesmidade e desconfiança, que regem a cadeia de custódia. Ao final, restou demonstrada a compatibilidade das referidas tecnologias com o processo penal brasileiro e com a cadeia de custódia, uma vez que as funcionalidades oferecidas pela inteligência artificial e pelo *Blockchain* dão uma maior efetividade ao contraditório e aos princípios da mesmidade e desconfiança, diante da maior transparência e integridade fornecida aos vestígios digitais.

Demonstrada a compatibilidade de aplicação da IA e do *Blockchain* na cadeia de custódia, passou-se a descrever algumas aplicações práticas das referidas tecnologias.

No tocante a IA, verificou-se que, entre as diversas formas de aprendizado de máquina, o aprendizado supervisionado e o aprendizado aprofundado seriam os mais adequados para as finalidades pretendidas pela cadeia de custódia, tendo em vista a possibilidade de se utilizar o aprendizado supervisionado para identificar vestígios digitais que sofreram algum tipo de adulteração no seu conteúdo, enquanto o aprendizado aprofundado serviria para identificar, em um vestígio digital, quais dados são relevantes para a investigação e quais não são excluindo-se da análise as informações impertinentes para a apuração do crime investigado, no intuito de proteger a privacidade dos investigados. Concluiu-se, assim, que a IA se mostra mais útil na fase inicial do procedimento da cadeia de custódia, isto é, durante a identificação e coleta do vestígio

nas aos arquivos que contenham determinadas palavras-chave relacionadas com os crimes investigados ao abrigo do mandado judicial, a fim de se obter um acesso progressivo ao conteúdo das informações e respeitar o direito à privacidade.

39 Yermack explica o funcionamento do Blockchain por meio do bitcoin: “to create a new block in the Bitcoin blockchain, the operator of a “node” on the network must bundle together transaction data, the hash code from the header of the prior block, the time stamp, and a further piece of data known as a “nonce.” The nonce is a random number with the property that, when added to the other information in a block, it generates a hash with a certain number of leading zeroes.⁴⁰ Once the fastest (or luckiest) miner finds a nonce and successfully completes a block with the required hash, network members will then verify and acknowledge the new block and begin working on its successor. A winning hash can only be discovered through trial-and-error, a computationally costly “proof of work” process that deters hackers from attempting to update the Blockchain with fraudulent data.” YERMACK, David. Corporate governance and Blockchains. Review of Finance, Oxford.

40 MOREIRA, Athur Salles de Paula; DELGADO, Camila Campos Baumgratz; SANTOS, Gabriel Gonçalves. Repensando a tecnologia block-

chain: por que nem tudo o que você leu até hoje era verdade? In: PARENTONI, Leonardo; MILAGRES, Marcelo de Oliveira; VAN DE GRAAF, Jeroen (Coords.). MOREIRA, Arthur Salles de Paula; CHAGAS, Ciro Costa; SANTANA, Mariana Damiani (Orgs). Direito, Tecnologia e Inovação, v. III: Aplicações Jurídicas de Blockchain. Expert Editora: Belo Horizonte, 2021, p. 31.

41 Matida ressalta a importância do registro da cadeia de custódia: “Não basta que se realize uma “custódia fática” a qual reflita o menor ou maior cuidado que agentes guardem a posse do objeto, individualmente, julguem ser devidos à sua preservação. Os novos dispositivos assentam que a confiabilidade dependerá do necessário oferecimento de um detalhado rastro, nos moldes definidos normativamente.” MATIDA, J. A cadeia de custódia é condição necessária para a redução dos riscos de condenações de inocentes. Revista da Defensoria Pública do Estado do Rio Grande do Sul, Porto Alegre, n. 27, 2021, p. 21. Disponível em: <https://revista.defensoria.rs.def.br/defensoria/article/view/269>. Acesso em: 26 jun. 2023.

42 WÜST, Karl; GERVAIS, Arthur. Do you need a Blockchain? 2018 crypto valley conference on Blockchain technology (CVCBT). IEEE, 2018, p. 47-48.

digital, diante da possibilidade de identificar e excluir, desde o princípio, vestígios digitais que tenham sofrido algum tipo de adulteração, impedindo que essas evidências cheguem ao processo.

Em relação ao *Blockchain*, constatou-se que suas qualidades, quais sejam, a verificabilidade pública, a transparência e a integridade vão ao encontro dos princípios que regem a cadeia de custódia, uma vez que possibilitam um registro fidedigno, seguro e imutável de todo o trajeto percorrido pelo vestígio digital desde sua identificação até a sua apresentação no processo, bem como a escrituração de todas as pessoas que tiveram contato com o vestígio, o motivo pelo qual tiveram o contato e todos os procedimentos ao qual o vestígio foi submetido. Além disso, o *Blockchain* pode controlar o acesso aos vestígios digitais, de modo que somente as partes autorizadas podem ter contato com os vestígios registrados na rede, vedando o acesso de pessoas não autorizadas. Dessa forma, chegou-se à conclusão de que todas as funcionalidades contidas no *Blockchain* fortalecem a confiabilidade dos vestígios digitais e aumentam o seu valor como prova no processo.

Destarte, todas as hipóteses apresentadas na introdução foram confirmadas ao final deste trabalho, que demonstrou que o uso em conjunto da IA e do *Blockchain* permite um procedimento mais preciso, transparente e seguro da cadeia de custódia, inexistindo qualquer óbice por parte do ordenamento jurídico brasileiro em relação ao emprego das referidas tecnologias.

Referências

- BADARÓ, Gustavo Henrique. *Processo penal*. 8. ed. São Paulo: Thomson Reuters Brasil, 2020.
- BRASIL. Decreto-Lei 3.689, de 03 de outubro de 1941. *Código de Processo Penal*. Brasília, DF: Presidência da República, 2020. Disponível em: http://www.planalto.gov.br/ccivil_03/decreto-lei/del3689.htm. Acesso em: 20 jun. 2023.
- BRASIL. Superior Tribunal de Justiça. Quinta Turma. Agravo Regimental em Habeas Corpus n.º 143.169/RJ. Agravante: R. L. S. M. Agravado: Ministério Público do Estado do Rio de Janeiro. Advogado: Marcelo Turbay Freiria. Rel. Min. Jesuíno Rissato. Brasília, 07 fev. 2023. Disponível em: <https://ww2.stj.jus.br/processo/pesquisa/>. Acesso em: 26 jun. 2023.
- CHATTERJEE, Marina. *Data Science vs Machine Learning and Artificial Intelligence: the difference explained*, 2023. Disponível em: <https://www.mygreatlearning.com/blog/difference-data-science-machine-learning-ai/>. Acesso em: 28 abr. 2023.

CHATTERJEE, Marina. *Deep Learning Tutorial: What it Means and what's the role of Deep Learning*, 2023. Disponível em: <https://www.mygreatlearning.com/blog/what-is-deep-learning/#one>. Acesso em: 17 jun. 2023.

DEZEM, Guilherme Madeira. *Curso de processo penal*. 8. ed. rev., atual. e ampl. São Paulo: Thomson Reuters Brasil, 2021.

KOTSIANTIS, Sotiris B. et al. *Supervised machine learning: A review of classification techniques. Emerging artificial intelligence applications in computer engineering*, v. 160, n. 1, p. 3-24, 2007.

LECUN, Yann; BENGIO, Yoshua; HINTON, Geoffrey. *Deep learning*. *Nature*, v. 521, n. 7553, p. 436-444, 2015.

LIMA, Gabriel Ribeiro de; NETO, Marcos Luiz dos Mares Guia. *Arbitragem em Blockchain: Juridicidade e perspectivas* In: PARENTONI, Leonardo; MILAGRES, Marcelo de Oliveira; VAN DE GRAAF, Jeroen (coords.). MOREIRA, Arthur Salles de Paula; CHAGAS, Ciro Costa; SANTANA, Mariana Damiani (orgs.). *Direito, Tecnologia e Inovação*, v. III: *Aplicações Jurídicas de Blockchain*. Expert Editora: Belo Horizonte, 2021.

LOPES JR., Aury. *Direito processual penal*. 16. ed. São Paulo: Saraiva Educação, 2019.

MAINI, Vishal. SABRI, Samer. *Machine Learning for humans*, 2017. Disponível em: <https://everything-computerscience.com/books/Machine%20Learning%20for%20Humans.pdf>. Acesso em: 28 abr. 2023.

MATIDA, J. *A cadeia de custódia é condição necessária para a redução dos riscos de condenações de inocentes*. *Revista da Defensoria Pública do Estado do Rio Grande do Sul*, Porto Alegre, n. 27, p. 17-26, 2021. Disponível em: <https://revista.defensoria.rs.def.br/defensoria/article/view/269>. Acesso em: 26 jun. 2023.

MOREIRA, Arthur Salles de Paula; DELGADO, Camila Campos Baumgratz; SANTOS, Gabriel Gonçalves. *Repensando a tecnologia Blockchain: por que nem tudo o que você leu até hoje era verdade?* In: PARENTONI, Leonardo; MILAGRES, Marcelo de Oliveira; VAN DE GRAAF, Jeroen (Coords.). MOREIRA, Arthur Salles de Paula; CHAGAS, Ciro Costa; SANTANA, Mariana Damiani (Orgs.). *Direito, Tecnologia e Inovação*, v. III: *Aplicações Jurídicas de Blockchain*. Expert Editora: Belo Horizonte, 2021.

OPENAI. *OpenAI's GPT-3: Language Models are Unsupervised Multitask Learners*. Disponível em: <https://openai.com/research/gpt-3/>. Acesso em: 28 abr. 2023.

PACHECO, César Augusto Rodrigues; PEREIRA, Natasha Sophie. *Deep learning conceitos e utilização nas diversas Áreas do conhecimento*. *Revista Ada Lovelace*, v. 2, p. 34-49, 2018.

PRADO, Geraldo. *A cadeia de custódia da prova no processo penal*. 2. ed. Rio de Janeiro: Marcial Pons, 2021.

SCHNEIER, Bruce. *Secret & Lies: Digital Security in a Networked World*. Indianapolis: Wiley Publishing, 2000.

TAKAHASHI, Tadao. *Sociedade da Informação no Brasil: o livro verde*. Brasília: Ministério da Ciência e Tecnologia, 2000.

U.S. DEPARTMENT OF JUSTICE OFFICE OF JUSTICE. *Electronic Crime Scene Investigation: A Guide for First Responders*. Second Edition. Programs National Institute of Justice. Disponível em: <https://www.ncjrs.gov/pdffiles1/nij/219941.pdf>. Acesso em: 26 jun. 2023.

43 Acerca da imutabilidade e integridade da rede Blockchain, Lima e Neto esclarecem que: "A integridade das informações é garantida por um sistema descentralizado de consenso, no qual os denominados mineradores verificam e adicionam transações ao livro razão. A criptografia assimétrica, o sistema distribuído de verificação e a imutabilidade da cadeia de blocos são algumas das principais características que tornam o Blockchain imutável e seguro, diferindo-o dos meios de transações tradicionais, que operam a partir de autoridades centralizadoras" LIMA, Gabriel Ribeiro de; NETO, Marcos Luiz dos Mares Guia. *Arbitragem em Blockchain: Juridicidade e perspectivas* In: PARENTONI, Leonardo; MILAGRES, Marcelo de Oliveira; VAN DE GRAAF, Jeroen (Coords.). MOREIRA, Arthur Salles de Paula; CHAGAS, Ciro Costa; SANTANA, Mariana Damiani (Orgs.). DIREITO, TECNOLOGIA E INOVAÇÃO – v. III: *Aplicações Jurídicas de Blockchain*. Expert Editora: Belo Horizonte, 2021, p. 110-111.

44 Segundo Urazán Bautista (2005 apud PRADO, 2021, p. 153), "a cadeia de custódia deve estar conformada pelo menor número de custódios possível; o menor número faz com que o elemento seja menos manipulado; a menor manipulação expõe menos; ao expor-se menos, se está protegendo o elemento, se lhe está defendendo." PRADO, Geraldo. *A cadeia de custódia da prova no processo penal*. 2.

ed. Rio de Janeiro: Marcial
Pons, 2021.

VAZ, Denise Provasi. *Provas digitais no processo penal: formulação do conceito, definição das características e sistematização do procedimento probatório*. 2012. 198 f. Tese (Doutorado em Direito) - Faculdade de Direito, Universidade de São Paulo, São Paulo, 2012. Disponível em: https://teses.usp.br/teses/disponiveis/2/2137/tde-28052013-153123/publico/Denise_Provasi_Vaz_tese_integral.pdf. Acesso em: 22 jun. 2023.

WÜST, Karl; GERVAIS, Arthur. *Do you need a Blockchain?* Crypto Valley Conference on Blockchain Technology (CVCBT), IEEE, p. 45-54, 2018.

YERMACK, David. *Corporate governance and Blockchains*. *Review of Finance*, Oxford, v. 21, p. 7-31, 2017.





■ Giselle Beiguelman, frames do vídeo generativo da série Flora Rebellis, versão 2024.
Imagens geradas com Inteligência Artificial (Style Gans)

■ Acesse o vídeo clicando na imagem





MONITORAMENTO DE INTEGRIDADE ESTRUTURAL UTILIZANDO INTELIGÊNCIA ARTIFICIAL:

Desafios, avanços e aplicações

DOI: <https://doi.org/10.35699/2965-6931.2023.47533>

VICTOR HIGINO MENEGUITTE ALVES*; VINICIUS ANTÔNIO MENEGUITTE ALVES**;
ALEXANDRE ABRAHÃO CURY***

RESUMO: Este artigo apresenta uma revisão sistemática e traz reflexões críticas acerca da utilização de técnicas de inteligência artificial para identificação de deterioração estrutural por meio de sinais de vibração (i.e., acelerações, deslocamentos, etc). Abordagens baseadas em aprendizado de máquina e aprendizado profundo são consideradas promissoras para aumentar a segurança e otimizar os cronogramas de manutenção preventiva. Entretanto, alguns autores reconhecem as preocupações decorrentes de métodos estritamente supervisionados, da natureza do tipo “caixa-preta” dos modelos e de suas interpretabilidades por operadores humanos. A contribuição deste trabalho consiste, portanto, em fornecer informações relevantes sobre o atual paradigma de detecção de danos, possibilitando previsões em tempo real, não destrutivas e confiáveis sobre a segurança da construção no âmbito da Indústria 4.0. Além disso, os desafios relacionados ao emprego de inteligência computacional para reconhecimento de padrões e tomada de decisão no monitoramento de anomalias estruturais são relatados e examinados em estudos de casos recentes.

PALAVRAS-CHAVE: detecção de danos estruturais; inteligência computacional; aprendizado de máquina; aprendizado profundo.

Structural integrity monitoring using artificial intelligence: Challenges, advances and applications

ABSTRACT: This paper presents a systematic review and brings critical reflections on the use of artificial intelligence techniques to identify structural deterioration through vibration signals (i.e., accelerations, displacements, etc.). Approaches based on machine learning and deep learning are considered promising for increasing safety and optimizing preventive maintenance schedules. However, some authors recognize concerns arising from strictly supervised methods, the “black box” nature of the models and their interpretability by human operators. Therefore, the contribution of this work is to provide relevant information about the current damage detection paradigm, enabling real-time, non-destructive, and reliable predictions about construction safety within the scope of Industry 4.0. Furthermore, challenges related to the use of computational intelligence for pattern recognition and decision-making in monitoring structural anomalies are reported and examined in recent case studies.

KEYWORDS: damage detection; constructions; computational Intelligence; machine learning; deep learning.

* Universidade Federal de Juiz de Fora/ ** Universidade Federal de Juiz de Fora/ *** Universidade Federal de Juiz de Fora

Introdução: antecedentes e novos desafios

A predição do estado atual de segurança de uma determinada estrutura é uma tarefa importante para manter o seu desempenho durante toda a sua vida útil. Entretanto, essa predição é altamente dependente das condições ambientais, cargas de uso, definições de projeto e muitos outros fatores, implicando em uma interação bastante complexa e frequentemente impraticável de serem analisadas conjuntamente. Inicialmente, a avaliação de estruturas surge por inspeções visuais humanas que podem ser inviáveis em locais de acesso remoto, além de serem uma análise subjetiva do engenheiro/técnico e ainda enviesadas por serem limitadas apenas à superfície do elemento construtivo, visto que danos internos à estrutura não poderiam ser identificados desta maneira.

Assim, os sistemas de monitoramento de integridade estrutural (SHM, do inglês *Structural Health Monitoring*) atraem muita atenção de pesquisadores em diversos domínios, como Engenharia Civil (Zhang; Lei, 2021), Mecânica (Wang; Lu; Yan., 2019), Ferroviária (Meixedo *et al.*, 2021) e Aeroespacial (Yuan, 2016), devido à sua abordagem não destrutiva e em tempo real que tem potencial para economizar custos com manutenção, visto que a periodicidade de reparos pode ser otimizada. Geralmente, as tarefas de SHM são divididas em quatro estágios, com níveis mais altos exigindo mais esforço para serem concluídos, a saber: (I) relatar a existência do dano, (II) localizar o dano, (III) quantificar o dano e (IV) determinar a vida estrutural restante (Rytter, 1993). A união desses níveis forma um sólido conjunto que sustenta a comunidade científica de SHM, fornecendo-lhes uma ferramenta poderosa para gerenciar as complexidades da identificação de danos.

Nesse contexto, a maioria dos métodos convencionais para SHM utiliza a análise modal, que se apoia na identificação de parâmetros modais (Moughy; Casas, 2017) como frequências naturais, taxas de amortecimento e modos de vibração. Contudo, esse processo de identificação modal é uma filtragem que pode levar à perda de informações, ocultando pequenas alterações estruturais que podem estar relacionadas a danos (Morales; Cury, 2018). De fato, a pesquisa conduzida por Morales e Cury (2018) evidenciou que, de modo geral, não é possível estabelecer uma relação clara entre as variações de temperatura e os efeitos de danos estruturais sobre as frequências natu-

rais de vibração de uma viga biapoiada testada em laboratório. Além disso, os métodos baseados em parâmetros modais assumem que a estrutura permanece no domínio elástico linear após o surgimento do dano, o que nem sempre corresponde à realidade (Cury; Crémona; Diday, 2010).

Uma outra abordagem consagrada, partindo de visão computacional, é discutida detalhadamente por Dong e Catbas (2021), a qual automatiza o processo de inspeção visual, mas mantém a desvantagem de só analisar o exterior dos elementos estruturais. Naquele trabalho, os autores destacam alguns desafios para se aplicar métodos baseados em visão computacional, como sombra, iluminação não-uniforme, mudanças no ângulo de inclinação e vibração da câmera. Dessa forma, anomalias ou processos de degradação podem ocorrer dentro dos componentes estruturais e permanecer ocultos à observação superficial. Descontinuidades, fissuras ou vazios no concreto podem estar entre esses problemas subjacentes (Tian *et al.*, 2021), assim como falhas ocultas, incluindo corrosão das armaduras embutidas, fibras fraturadas ou fissuras de delaminação em materiais compostos (Daneshjoo; Shokrieh; Fakoor, 2018). Por outro lado, o monitoramento baseado em vibração avalia a integridade das estruturas por meio de dados dinâmicos coletados por sensores posicionados nas próprias estruturas. Essa última estratégia será o enfoque principal no presente texto e é extremamente útil quando se torna difícil ou impraticável identificar problemas estruturais apenas visualmente.

Recentemente, abordagens têm sido desenvolvidas para explorar o potencial do uso de técnicas de inteligência artificial (IA) na detecção de degradação estrutural. Dentre essas ferramentas, destacam-se as Redes Neurais Artificiais (ANN, do inglês *Artificial Neural Networks*), Máquinas de Vetor de Suporte (SVM, do inglês *Support Vector Machine*), Conjunto de Árvores de Decisão (DT, do inglês *Decision Tree*), Redes Neurais Convolucionais (CNN, do inglês *Convolutional Neural Network*), dentre outras (AVCI *et al.*, 2021). Um sistema padrão de detecção de danos é mostrado na Figura 1, sendo composto por dois elementos principais: o *hardware* e o *software*. Na parte de *hardware* (Figura 1(b)), sensores (acelerômetros) distribuídos pela estrutura (Figura 1(a)) são utilizados para coletar as respostas dinâmicas através de um sistema de aquisição, podendo estes sinais serem gerados por excitação ambiental ou forçada. Já o elemento de *software* (Figura 1(c)) fica responsável pelo pré-processamento dos dados,

além do pós-processamento que pode se dar a partir de um algoritmo definido para reconhecimento de padrões, realizar a detecção, localização e/ou quantificação do dano. Note-se que nada impede a execução simultânea das quatro tarefas (I a IV) pelo mesmo método. Entretanto, esse problema se torna sucessivamente mais difícil. Assim, é comum que alguns autores se limitem a investigar uma tarefa específica, isto é, verificar apenas a existência do dano (equivalente ao nível I), por exemplo.

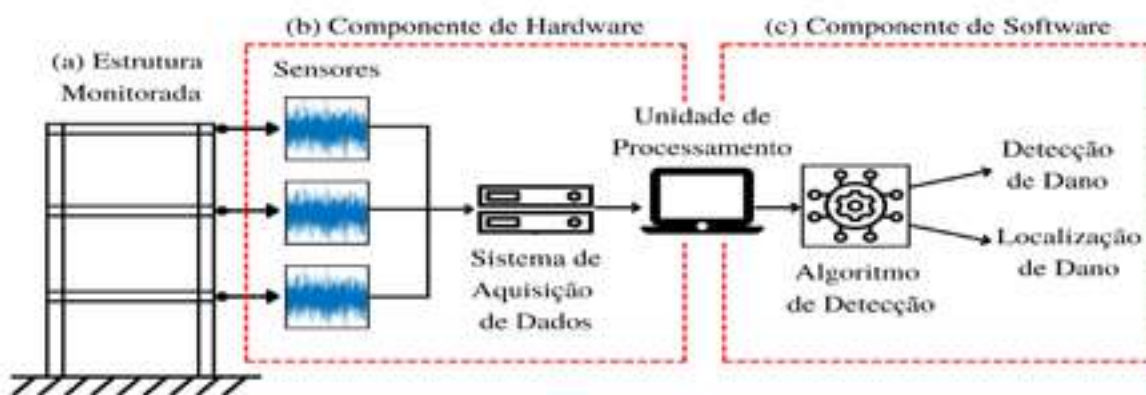


Figura 1 – Principais elementos de sistemas de detecção de danos estruturais.

Fonte: Elaboração própria, baseada e adaptada da obra de Avci *et al.* (2021).

Assim, desenvolvimentos recentes na tecnologia de sensores tendem a melhorar a confiabilidade e oferecer preços cada vez mais acessíveis para uso em aplicações de grande escala. Isso se torna uma conquista notável, pois a localização se torna mais precisa quanto mais sensores estão dispersos ao longo de uma estrutura. Ademais, os vários avanços na tecnologia de *chips*, juntamente com o aumento na capacidade de processamento, tornaram possível empregar aspectos da Indústria 4.0 ao monitoramento de estruturas. Como consequência, o processamento de sinais é um tópico que está em constante expansão, além de ser uma ferramenta fundamental para as tecnologias SHM, como resultado dos recentes desenvolvimentos em IA, aprendizado de máquina (ML, do inglês *Machine Learning*) e aprendizado profundo (DL, do inglês *Deep Learning*).

Nessa perspectiva, o objetivo do presente trabalho é fornecer uma visão abrangente dos avanços e desafios no monitoramento estrutural por meio de dados dinâmicos,

destacando o progresso alcançado nesse campo na Indústria 4.0. Desta forma, esta revisão sistemática se destaca em meio às já existentes por comparar as técnicas mais recentes de SHM baseadas em ciência de dados, ML e DL quanto ao nível na escala de Rytter (1997) das tarefas que executam. Outro aspecto relevante e original deste trabalho é a discussão crítica do caráter supervisionado ou não-supervisionado dos algoritmos, as implicações disso em uma situação real de monitoramento, além das metodologias que diversos autores têm trazido para solucionar esses problemas. Esta pesquisa também faz um levantamento dos principais dados de *benchmark*, ou seja, de estruturas ensaiadas experimentalmente que possuem maior recorrência na literatura por se tratarem de estudos de caso bem consolidados e com dados abertos à comunidade científica. Assim, a lista de bases de dados recomendada possibilita a comparação facilitada entre os métodos do estado da arte. Por fim, traremos reflexões éticas acerca da interpretabilidade dos modelos de inteligência computacional, os quais tendem a atuar como “caixas-pretas”, seguido de uma discussão sobre técnicas disponíveis para reduzir esse tipo de comportamento. Desse modo, o presente artigo servirá, também, como um manual introdutório de boas práticas no âmbito do campo de estudo do SHM baseado em IA.

Metodologia

Este artigo foi desenvolvido a partir de pesquisas bibliográficas. Foram considerados apenas artigos entre os anos de 2017 e 2023, visando abranger os estudos mais atuais sobre o tema. Vale destacar que essa compilação não pretende ser uma abordagem exaustiva de todos os métodos de SHM com dados dinâmicos desenvolvidos até o momento, mas sim uma coleção criteriosamente selecionada de artigos revisados por pares, oferecendo uma visão panorâmica do estado da arte atual. Utilizando uma estratégia de busca sistemática, foram identificados e incluídos nesta compilação os trabalhos mais relevantes, provenientes de reconhecidas bases acadêmicas como Web of Science, Science Direct, Wiley Online Library e Google Acadêmico, com um enfoque em artigos de revistas mais expressivas do campo de estudo de SHM. A meta é

propiciar uma análise panorâmica do progresso alcançado nesse campo, reconhecendo os avanços e desafios que norteiam o monitoramento estrutural através de dados de vibração na era da Indústria 4.0. Os termos e conectivos booleanos usados para combiná-los na busca foram: (“Vibration”) E (“Damage”) E (“Detection” OU “Localization”) E (“SHM”) E (“AI” OU “ML” OU “DL”). Os termos são inseridos em inglês devido à sua presença majoritária nas revistas mais relevantes.

Posteriormente, foram avaliados os títulos, resumos e palavras-chave, a fim de excluir trabalhos que não tivessem a ver com abordagens de uso de algoritmos de inteligência computacional aplicados ao monitoramento de estruturas. Dos documentos restantes, 16 foram selecionados, lidos na íntegra e julgados pelos autores como sendo os estudos que melhor representam, de forma abrangente, o estado da arte do tema. Após a leitura, esses 16 artigos tiveram seus dados catalogados e as suas descobertas e resultados apresentados e discutidos. Além disso, esse trabalho é baseado também em artigos do tipo revisão da literatura com o intuito de se verificar o que é produzido de característica mais inovadora na área de detecção de danos estruturais com base em dados de vibração com IA. Portanto, para detalhes complementares sobre o tema, sugere-se os trabalhos de Avci et al (2021); Azimi, Eslamlou e Pekcan (2020). Uma visão geral da metodologia utilizada neste trabalho é ilustrada na Figura 2.

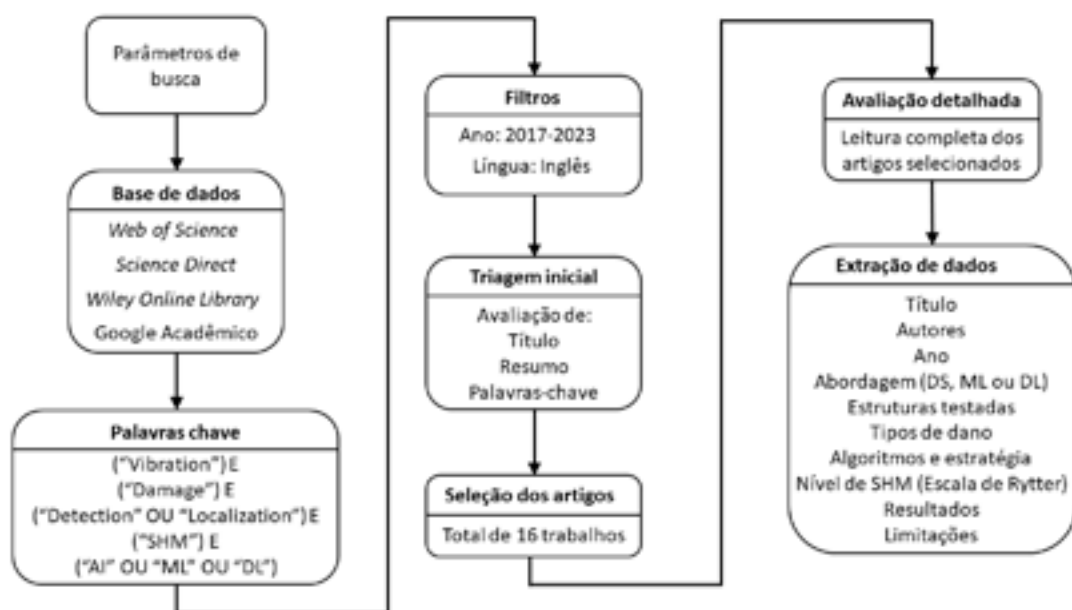


Figura 2 – Fluxograma da revisão sistemática proposta.

Fonte: Elaboração própria.

Técnicas de identificação de anomalias e seus resultados

Essa seção apresenta alguns dos resultados de trabalhos mais recentes na área de pesquisa de detecção de danos estruturais baseados em dados de vibração. Dessa maneira, são apresentadas abordagens através de ciência de dados e aprendizado de máquina e aprendizado profundo, explorando as suas vantagens e limitações.

Detecção de danos com ciência de dados (ds, do inglês data science) e aprendizado de máquina (ml)

O trabalho de Alves e Cury (2021) utiliza, fundamentalmente, conceitos de ciência de dados e estatística para localizar danos estruturais. A partir de dados de vibração coletados no estado intacto (referência) e do estado atual (desconhecido) da estrutura, extrai-se as características no domínio do tempo, frequência e quefrência. Depois de extraídas 35 características dos sinais brutos de aceleração, é fundamental construir um índice que reflita objetivamente o(s) local(is) do dano. Como resultado, a técnica sugerida começa determinando intervalos percentis para cada uma das características associadas ao estado intacto da estrutura. Em seguida, é feita uma análise de *outliers* (valores aberrantes, ou anormais) comparando-se as características do estado desconhecido aos intervalos criados a partir da estrutura de referência (saudável). Como resultado, um índice de dano é calculado de modo que quanto maior o número de características discrepantes em um determinado local (sensor), maior é o índice de dano, correspondendo à maior probabilidade de dano. Note que esse processo de comparação dos sinais intactos com sinais desconhecidos da estrutura (possivelmente danificada) é uma ideia recorrente entre os métodos da literatura. A Figura 3 ilustra esse esquema típico.

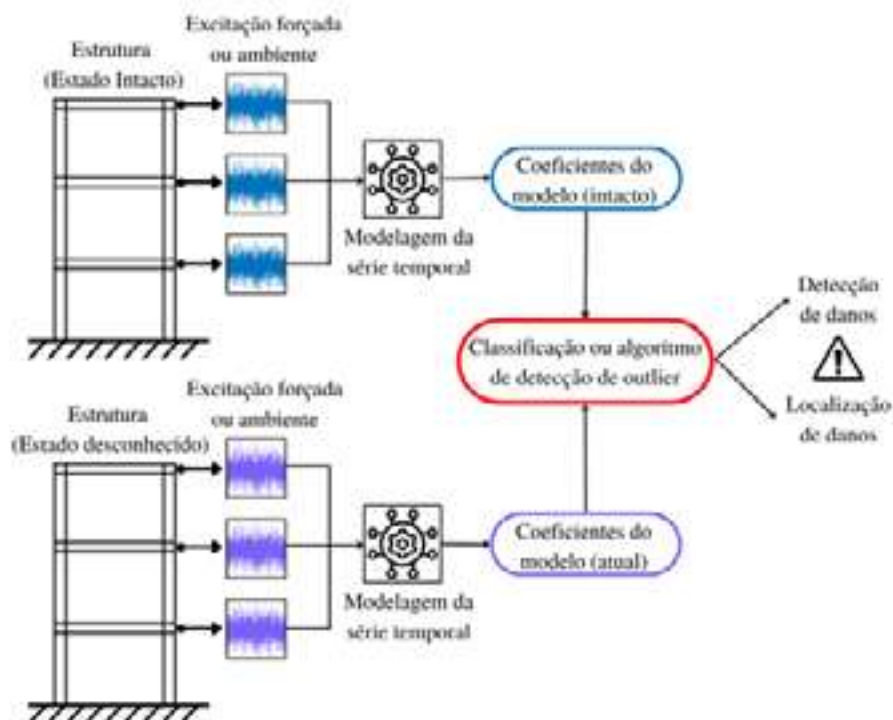


Figura 3 – Esquema típico de detecção de danos com inteligência artificial a partir de sinais dinâmicos.

Fonte: Elaboração própria, baseada e adaptada da obra de Avci *et al.* (2021).

Essa metodologia foi melhorada e totalmente automatizada em Alves e Cury (2023). Dessa vez, entretanto, as características extraídas foram posteriormente filtradas por uma técnica de seleção de recursos não supervisionada denominada InfFS, com base em sua relevância e não redundância. Além disso, um novo índice de dano ponderado foi empregado para investigar a existência de um grau de importância no conjunto de características selecionadas. Apesar de ser completamente não supervisionado e automatizado, o método alcançou resultados comparáveis a abordagens tradicionais e até recentes técnicas de aprendizado profundo em todas as aplicações testadas. Em um experimento de viga biapoada, o método revelou-se promissor em detectar danos extremamente sutis, como pequenos furos, que apresentaram variações inferiores a 1% da primeira frequência natural. Essa capacidade demonstra a sua viabilidade na identificação precoce de danos em cenários em que abordagens baseadas em análise

modal seriam ineficazes. Outrossim, o método igualmente evidenciou a sua sensibilidade em detectar danos consideravelmente graves como, por exemplo, uma perda de rigidez de 51,7% em 4 pilares de aço de um pórtico de pequena escala.

Em ambos os casos, o modelo de detecção foi criado apenas com os dados intactos (processo não-supervisionado). Entretanto há casos em que o esquema utilizado é o de um treinamento supervisionado. Assim, traz-se à tona a questão de como os dados danificados seriam gerados para treinar os modelos, haja vista que seria contraditório danificar propositalmente a estrutura.

Uma estratégia supervisionada do estudo de Gui et al (2017), investigou a viabilidade de se utilizar máquinas de vetores de suporte (SVM) associadas a técnicas de otimização para extração aprimorada de características e parâmetros de otimização para detecção de danos. Dois tipos de métodos de extração de características foram avaliados, incluindo parâmetros de modelos autorregressivos (AR, do inglês *AutoRegressive*) e erros residuais de parâmetros estatísticos. Assim, foram testadas três técnicas de otimização: método de busca em grade (GS, do inglês *Grid Search*), otimização de enxame de partículas (PSO, do inglês *Particle Swarm Optimization*) e algoritmo genético (GA, do inglês *Genetic Algorithm*) para determinar, efetivamente, os parâmetros no SVM.

Assim, fica evidente a discussão da existência de possíveis níveis diferentes de importância entre as características usadas como entrada. Sob essa ótica, o trabalho de Ghiasi, Ghasemi e Chan (2021) realizou uma extração de características usando pacotes de energia relativa *wavelet* (WPRE) e sugeriu um método de Seleção de Características (FS, do inglês *Feature Extration*) para escolher as características mais relevantes e não-redundantes. Para a classificação foi utilizada uma versão melhorada de SVM. A cada etapa, foram realizadas comparações com outras métricas mostrando maior precisão na detecção da existência de danos. O método foi testado em seis estruturas com acurácias de 79,8% até 99,7%.

Para mitigar a influência de um fator ambiental externo, o estudo de Finotti *et al.* (2017) utilizou as frequências naturais e medições de temperatura da estrutura como entrada para o método SVM. Para verificar a metodologia proposta, foram utilizados dados experimentais da Torre Gabbia, na Itália, que incluem registros do comportamento estrutural antes, durante e após um evento sísmico. Ao analisar os resultados da aplicação na torre, verificou-se que a técnica baseada nas frequências naturais foi

eficiente em detectar mudanças estruturais, com resultados em concordância com a situação real da torre, permitindo identificar claramente os danos causados pelo evento sísmico. Entretanto, revela o desafio de se adquirir os dados de treinamento dos modelos em diferentes cenários estruturais.

Em uma tentativa de solucionar a questão dos dados danificados para treino, sob uma perspectiva baseada em modelos, Mariniello *et al.* (2021) produziu uma biblioteca de respostas estruturais “patológicas” e “não patológicas” virtuais usando modelos em elementos finitos calibrados por testes de laboratório e aplicou-a a uma metodologia de vibração atrelada a um método de aprendizado baseada em árvores de decisão. O método foi utilizado para identificar respostas anormais, indicar a localização e a severidade de dano. Sua abordagem foi capaz de localizar danos únicos, mas mostrou precisão reduzida ao lidar com dois ou mais elementos danificados. Em geral, a metodologia proposta classificou os danos aprendendo as propriedades de um conjunto de dados simulados, expondo a desvantagem de que, se a modelagem do sistema estrutural não for realista o suficiente, a abordagem se torna inválida.

Assim, visando minimizar essas limitações, o trabalho de Zacharakis e Giagopoulos (2022) apresentou uma técnica de detecção de danos utilizando método dos elementos finitos (FEM, do inglês *Finite Element Method*), algoritmos meta-heurísticos e medições experimentais. O objetivo principal era facilitar uma estratégia de inspeção baseada em vibração de estruturas explorando informação de monitoramento. Como ponto de partida, desenvolveu-se um modelo FE a partir de um algoritmo de Otimização por Enxame de Partículas (PSO, do inglês *Particle Swarm Optimization*) que pode descrever a resposta dinâmica da estrutura saudável com precisão. O modelo FE otimizado, juntamente com medições experimentais de vibração da estrutura, compõem as entradas da detecção de danos, enquanto a saída é um modelo FE danificado que aproxima a resposta dinâmica da estrutura danificada do mundo real. Para tanto, uma área danificada é inserida no modelo FE ótimo, alterando os parâmetros de rigidez e massa para aproximar o efeito do dano físico. Após a convergência do procedimento de otimização, a área inserida destaca a região danificada da estrutura. Foram testadas duas estruturas: o chassi de um veículo e uma barra de material compósito. O maior erro na modelagem das estruturas testadas foram 2,38% e 2,14%, respectivamente. Ademais, os cenários de danos causaram mudanças na resposta dinâmica das estru-

turas físicas de 2,5% até 5,39%. A abordagem proposta foi capaz de encontrar a área danificada no total de 4 cenários estudados.

Com base em uma coletânea representativa de 22 casos de uso da ciência de dados em SHM, o livro de Cury *et al.* (2022) expõe diversos fatores críticos que impactam o desempenho dos algoritmos de identificação de danos em sistemas estruturais de grande escala. Esses desafios também são evidenciados nos demais estudos apresentados nesta seção. Primeiramente, no âmbito do *Big Data*, a alta dimensionalidade dos parâmetros monitorados representa um desafio na análise e processamento efetivo da vasta quantidade de dados envolvidos. Além disso, fatores ambientais e operacionais, como temperatura, umidade e tráfego, podem introduzir variabilidade e incerteza nas medições, dificultando ainda mais o processo de identificação. A complexidade inerente das grandes estruturas também contribui para a dificuldade em detectar e localizar danos com precisão. Além disso, garantir a confiabilidade dos dados medidos é fundamental para obter resultados precisos e confiáveis. Outro problema significativo é a sensibilidade limitada das respostas estruturais globais a danos locais, o que pode dificultar a detecção precoce de anomalias sutis. Por fim, entende-se que o monitoramento eficaz da integridade estrutural requer a integração do conhecimento físico e de engenharia em algoritmos de aprendizado de máquina, a fim de possibilitar um processo de dados para decisões eficientes.

Detecção de danos com aprendizado profundo (dl)

Pesquisas recentes na literatura utilizam aprendizado profundo (DL) para detectar danos estruturais, entendendo esse problema como uma tarefa de classificação de séries temporais (Sony *et al.*, 2022). Nessa perspectiva, as Redes Neurais Convolucionais (CNN) são as mais utilizadas, visto que se tornaram o estado da arte de DL para reconhecimento de padrões em imagens (Ciresan *et al.*, 2010). Entretanto, para o uso em sinais de vibração, adota-se comumente Redes Neurais Convolucionais de uma dimensão (1D CNN), visto que, neste caso, trata-se de vetores (sinais), e não matrizes (imagens). A vantagem dessa abordagem é a sua capacidade de extrair automaticamente as características sensíveis a danos dos sinais de aceleração através do treinamento de camadas de neurônios subsequentes, enquanto outros métodos convencionais de mineração de dados usam características escolhidas à mão, as quais podem ser uma

opção não otimizada para uma estrutura em particular ou falhar em alcançar o mesmo nível de desempenho em outras estruturas.

Desse modo, a estratégia desenvolvida no trabalho de Abdeljaber *et al.* (2017) consistiu em uma metodologia baseada em 1D CNN para a localização de danos em juntas de uma estrutura de aço testada em laboratório, como mostrado na Figura 4. Esse estudo possui grande importância por introduzir um importante *benchmark* chamado Qatar University Grandstand Simulator (QUGS), o qual possibilitou diversas aplicações de DL no campo do SHM desde então. Basicamente, o dano consiste em um afrouxamento dos parafusos que conectam a união entre vigas, simulando uma deterioração estrutural. O objetivo do algoritmo é identificar corretamente a posição de quaisquer articulações danificadas. Cada uma das 30 articulações equipadas com acelerômetros precisa de sua própria 1D CNN, que deve ser projetada e treinada especificamente para isso como parte da metodologia sugerida. Cada CNN_{*i*} tem a tarefa de avaliar a saúde de uma única articulação *i* usando apenas os dados brutos de aceleração registrados nessa articulação. Portanto, o método de detecção de danos proposto é considerado descentralizado, pois a rede de uma articulação não tem o auxílio das demais CNNs.

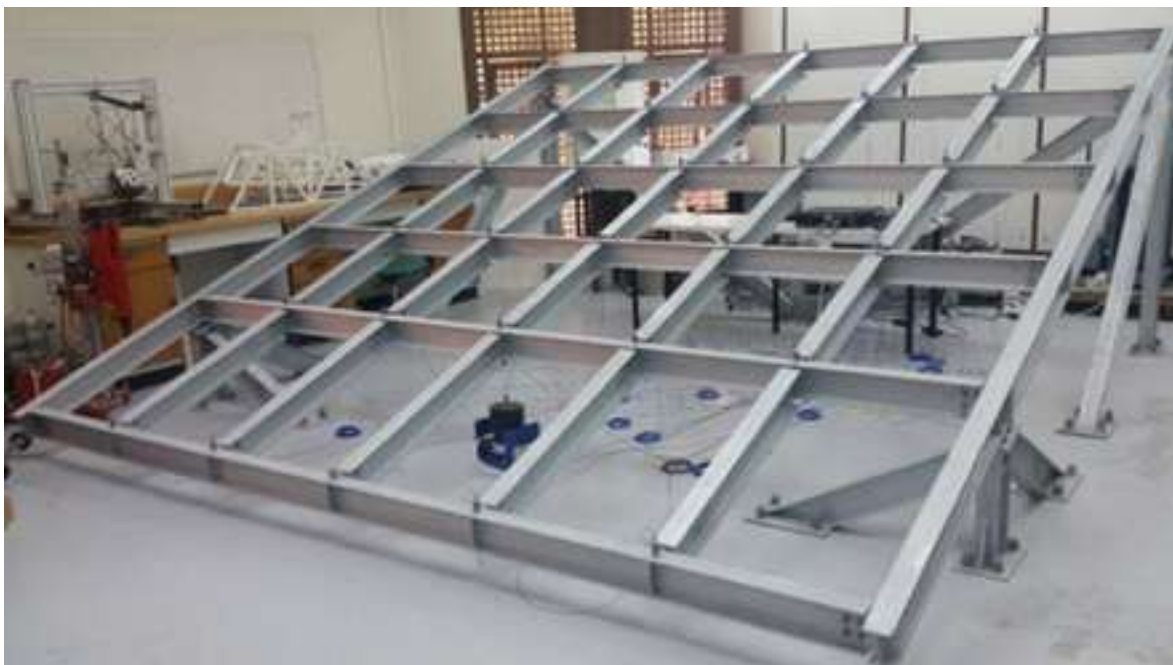


Figura 4 – Vista da estrutura de aço: QUGS.

Fonte: Abdeljaber et al (2017).

O conjunto de dados para treino foi definido de modo que os dados intactos de uma determinada junta contenham os sinais medidos, enquanto as outras juntas estavam intactas e/ou também danificadas, de modo a minimizar a influência do efeito do dano de uma junta sobre as outras no momento da classificação.

O teste da metodologia é iniciado induzindo dano em um ou mais locais (ou ainda, mantendo intacto) e aplicando um vibrador com excitações aleatórias. Então, os sinais de aceleração de cada junta são medidos e, em seguida, cada sinal é dividido em amostras de intervalos menores. Posteriormente, as amostras são normalizadas entre -1 e 1 e usadas para alimentar a sua CNN correspondente de cada junta. Assim, para cada junta, a probabilidade de dano é dada por $\frac{C}{N}$, onde C é o número total de amostras de intervalos de sinais processados pela CNN e N é o número de amostras classificadas como danificadas. Note que o resultado final é ponderado pelos resultados intermediários de cada amostra para maior representatividade dos dados. Desse modo, é esperado que o valor de $\frac{C}{N}$ seja mais próximo de 0 em juntas intactas e mais próximo de 1 em juntas danificadas. Alguns dos resultados desse método são ilustrados na Figura 5.

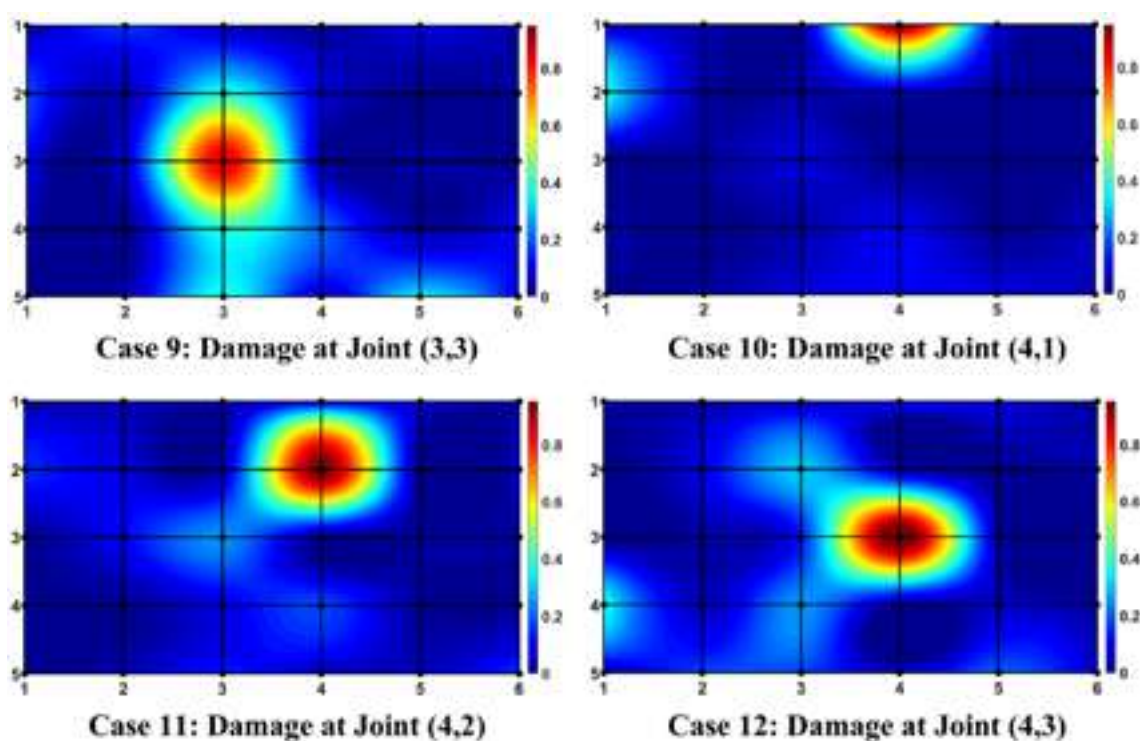


Figura 5 – Parte dos resultados de localização usando CNN.

Fonte: Abdeljaber et al. (2017).

A fim de avaliar o desempenho das 30 CNNs que foram produzidas, o algoritmo foi testado em 24 cenários estruturais (1 caso intacto, 18 casos de dano único e 5 casos de dano duplo). O erro médio de classificação das 30 CNNs foi determinado como sendo de apenas 0,54% nos dados treinados. Os testes demonstram um excelente desempenho para SHM em tempo real e detecção de danos estruturais.

Utilizando a mesma base de dados QUGS e adicionalmente a de uma ponte em escala real, o trabalho de Sony *et al.* (2022) propôs uma arquitetura de rede neural que utiliza camadas de *Long Short-Term Memory* (LSTM). Estas são redes com *loops*, permitindo que as informações persistam. Dessa maneira, a técnica de DL é treinada e aplicada com excelente precisão para classificar sinais de séries temporais em níveis de danos múltiplos e em multiclass. Por comparação, naquela pesquisa, a implementação da nova abordagem por LSTM apresentou resultados de localização de dano superiores a uma CNN equivalente.

Os métodos de detecção de anomalias possuem desafios referentes à operação sob condições variáveis. A maioria dos métodos existentes é baseada em uma suposição estacionária e constante, podendo ser levada ao erro devido a efeitos externos ou a mudanças de utilização das estruturas. Com isso, Luo *et al.* (2018) propuseram um novo método que permite a detecção de falhas precoces em condições de trabalho variáveis no tempo utilizando DL que utiliza camadas de *Stacked Auto-Encoders* (AE) e *Back Propagation Neural Network* (BPNN). Um modelo foi desenvolvido com o propósito de automatizar a seleção das respostas de impulso a partir de grandes conjuntos de dados de vibração coletados durante o funcionamento em longo prazo. Com base na análise da similaridade das propriedades dinâmicas, um índice de integridade foi construído para indicar o processo lento e gradual de deterioração da estrutura estudada. O trabalho validou a metodologia em um estudo experimental envolvendo uma máquina de usinagem por controle numérico computadorizado (CNC), que não tem um padrão de funcionamento constante, permitindo a adequada validação da metodologia proposta.

De maneira semelhante, Finotti *et al.* (2021) explorou uma proposta híbrida de uso de um *Sparse Auto-Encoder* (SAE) associado ao SVM para caracterizar os sinais de vibração, dessa vez com o objetivo de detectar anomalias em estruturas de engenharia civil. O SAE foi empregado para extrair características de maneira não-supervisionada diretamente das acelerações no domínio do tempo, e essas informações foram utiliza-

das como entradas para treinar um modelo supervisionado de classificação baseado no tradicional SVM. Inicialmente, a metodologia proposta foi testada em respostas estruturais de um modelo numérico de uma viga simplesmente apoiada, sujeita a diferentes níveis de dano. Em seguida, para verificar a eficácia do SAE/SVM em uma estrutura real, foram analisados dados experimentais do viaduto Várzea Nova em João Pessoa, Brasil. Neste estudo, foram realizados testes dinâmicos no viaduto antes e depois de um procedimento de reforço. O método alcançou mais de 99% de acurácia em ambos os casos.

O estudo de Alazzawi e Wang (2022) propôs uma Rede Residual Profunda (DRN, do inglês *Deep Residual Network*) para extrair e aprender automaticamente as características dos sinais brutos de aceleração. Um algoritmo de otimização Bayesiano foi utilizado para melhorar a eficiência dessa rede neural tanto em questão de precisão, quanto na escolha de parâmetros. O método sugerido foi capaz de efetuar detectar, localizar e quantificar o dano. Aplicações numéricas e experimentais foram realizadas em dois pórticos e uma ponte, alcançando precisão de 100% nos dados de teste empregados.

Particularmente, os métodos DL funcionam excepcionalmente bem, mas, como limitação, requerem uma grande quantidade de dados para operar. Para contornar esse problema, o estudo recente de Chamangard et al (2022) propôs uma técnica baseada em transferência de aprendizado (TL, do inglês *Transfer Learning*) para lidar com essa questão. O trabalho foi dividido em duas etapas. A primeira etapa tenta criar uma CNN compacta que, quando recebe dados suficientes de um modelo numérico, pode identificar com precisão os danos estruturais. A segunda parte examina a questão dos dados de treinamento em quantidades insuficientes, i.e., inadequados ao SHM de estruturas experimentais e do mundo real. Para resolver esse problema, o TL é empregado. A rede neural final obtém todo o conhecimento de uma CNN compacta treinada previamente em outra estrutura, que é usada como fonte. Assim, a informação extra obtida por esta rede pode ser utilizada para compensar a falta de dados. Ou seja, quando se inicia a aplicação em uma nova estrutura, o uso de TL tende a diminuir o volume de dados de treino necessário. O método foi avaliado somente quanto à classificação da existência ou não de dano (SHM nível I). A CNN compacta pode atingir 100% de precisão quando os dados estão disponíveis para treinamento. Além disso, para o caso de dados

insuficientes, o uso de uma rede compacta, bem como o aprendizado de transferência, proporciona melhorias consideráveis (cerca de 95%) na precisão da detecção de danos.

Uma alternativa ao grande volume de dados de treino foi elaborada na pesquisa de Luleci, Catbas e Avci (2023a) baseada nas Redes Adversariais Generativas (GAN, do inglês *Generative Adversarial Network*) Convolucionais Profundas de Wasserstein Unidimensionais usando Penalidade de Gradiente, a qual foi denominada 1-D WDCGAN-GP, para gerar dados sintéticos extras de aceleração do estado danificado da estrutura, ou seja, dados de dano que não foram adquiridos experimentalmente. A estratégia consiste fundamentalmente em duas redes neurais: uma chamada de geradora, que cria sinais de vibração, enquanto a outra, a discriminadora, avalia a sua autenticidade. Durante certo momento do aprendizado, a geradora passa a criar réplicas com comportamentos indistinguíveis das originais. A partir disso, os autores validaram os conjuntos de dados gerados com métodos quantitativos e qualitativos.

Em um avanço mais recente, Luleci, Catbas e Avci (2023b) empregaram uma variante inovadora do modelo GAN, o 1-D CycleWDCGAN-GP. Este modelo não apenas aprende a gerar amostras de dados semelhantes dentro de um único domínio, mas também domina o mapeamento complexo entre dois domínios de dados – não danificados e danificados. Um aspecto fascinante desse modelo é a sua capacidade de obter respostas “saúdáveis” de uma estrutura e traduzi-las em respostas danificadas, preenchendo a lacuna entre os dois domínios. Embora as comparações no domínio do tempo tenham demonstrado excelente desempenho, há espaço para aprimoramento adicional nas comparações no domínio da frequência, possivelmente por meio de treinamento mais diversificado. Para realizar essa proposta, os pesquisadores compilaram uma coleção de conjuntos de dados de resposta dinâmica de uma estrutura de aço, abrangendo conjuntos de dados danificados (afrouxamento de parafusos da junta) e não danificados, para treinar o modelo variante CycleGAN. Posteriormente, eles empregaram o conjunto de dados não danificado para tradução no domínio danificado e vice-versa. Os indicadores de desempenho usados no processo de avaliação garantiram que os resultados gerados em ambos os domínios se assemelhavam aos domínios verdadeiros, afirmando a eficácia do modelo na geração de dados autênticos.

Apesar de seus esforços, o resultado da tradução de domínio das juntas de teste foi limitado, pois apesar de conseguir traduzir o domínio para a junta treinada, não tinha

o mesmo sucesso com os domínios das outras juntas. Assim, em um estudo inovador, Luceli, Avci e Catbas (2023c) investigaram uma versão melhorada do modelo CycleGAN para obter maior precisão. Algumas das melhorias envolveram: um procedimento de treinamento mais abrangente para aumentar o conhecimento de domínio do modelo e a função de perda total do gerador foi complementada com uma nova perda baseada no domínio de frequência para melhor capturar o conteúdo de frequência dos dados. Este estudo inovador abriu possibilidades empolgantes para o monitoramento de integridade estrutural, com aplicações potenciais na previsão e simulação de respostas estruturais sob condições variadas. A tradução bem-sucedida de respostas de domínios saudáveis para domínios não saudáveis pode desempenhar um papel fundamental no avanço da compreensão do comportamento estrutural e, finalmente, no aprimoramento das medidas de segurança em infraestruturas críticas.

Discussão da interpretabilidade dos modelos

Conforme observado, as técnicas de ML e DL têm sido aplicadas com sucesso em diversos casos reais. No entanto, os algoritmos adotados nesses estudos para estruturas de monitoramento de integridade com base em vibrações (VSHM, do inglês *Vibration Structural Health Monitoring*) são frequentemente chamados de modelos “caixas-pretas” por oferecerem poucas informações sobre o processo de tomada de decisão (ARRIETA *et al.*, 2020). A falta de interpretabilidade e compreensão sobre como esses modelos detectam alterações na integridade estrutural e como utilizam as características obtidas dos dados para fazer previsões é uma das razões que dificultam a adoção generalizada dessas técnicas em aplicações industriais. Possibilitar a interpretabilidade em metodologias de VSHM baseadas em ML/DL pode fornecer evidências importantes para uma tomada de decisão confiável em aplicações de monitoramento de integridade estrutural. Em particular, a capacidade de distinguir entre anomalias causadas pela influência de variáveis operacionais externas ou danos reais na estrutura pode ajudar a reduzir falsos alarmes e aprimorar a precisão geral do sistema.

Segundo Arrieta *et al.* (2020), para definir os modelos de IA explicáveis, surgiram dois grupos principais de abordagens: (a) modelos transparentes, que oferecem explicabilidade direta por meio do próprio modelo ou de sua estrutura, incluindo árvores de decisão, modelos Bayesianos, modelos para k-vizinhos mais próximos e modelos de

regressão logística ou linear. Alguns destes, já mencionados em aplicações no presente artigo. (b) Explicabilidade pós-hoc, dividida em métodos agnósticos ao modelo (*model-agnostic* ou generalistas) e métodos específicos do modelo (*model-specific*), permite a explicabilidade subsequente dos modelos de ML. Métodos agnósticos ao modelo oferecem explicações usando ferramentas gerais aplicáveis a uma ampla variedade de algoritmos de ML. Por outro lado, os métodos específicos do modelo concentram-se em fornecer explicações para resultados obtidos por algoritmos de ML específicos, limitados a alguns algoritmos como redes neurais artificiais e máquinas de vetores de suporte.

Sob essa conjuntura, Amin *et al.* (2023) enfrentaram o desafio de interpretar modelos CNN “caixas-pretas” no contexto do monitoramento de condições para turbinas eólicas sob diversas condições de carga, especialmente cenários de vento turbulento. Para lançar luz sobre as previsões feitas pelo classificador, eles empregaram técnicas de explicações agnósticas de modelos interpretáveis locais (*Local Interpretable Model-Agnostic Explanations* - LIME). Ao fazer isso, eles desenvolveram um classificador interpretável baseado em física, oferecendo informações valiosas sobre o raciocínio por trás das decisões do modelo. A pesquisa se concentrou na detecção de falhas nos componentes rotativos da caixa de engrenagens, que deixam assinaturas de vibração distintas mensuráveis por meio de acelerômetros. Em seus experimentos, as falhas foram detectadas com sucesso.

O estudo de Brusa *et al.* (2023) usou uma técnica de interpretabilidade chamada *SHapley Additive exPlanation* (SHAP) com o objetivo para identificar as características mais relevantes na detecção e classificação de falhas em programas de monitoramento de condições para máquinas rotativas. Os pesquisadores investigaram especificamente rolamentos de tamanho médio de interesse industrial, coletando dados de vibração para diferentes estados de saúde em um equipamento de teste de rolamentos industriais disponível no Laboratório de Engenharia Mecânica do Politécnico de Turim. Os modelos de diagnóstico SVM e kNN (do inglês, *k-Nearest Neighbour*) foram explicados utilizando o SHAP e alcançaram acurácias superiores a 98,5% para ambos os modelos, ao utilizá-lo como critério para seleção das características.

As técnicas baseadas em IA discutidas neste artigo têm as suas características resumidas no Quadro 1. Ademais, entende-se como uma limitação geral dos métodos de detecção de dano baseados em vibração a disposição dos sensores instalados na

estrutura. Dessa maneira, se o dano ocorrer em um local afastado dos sensores, não seria possível detectá-lo. Entretanto, essa desvantagem se torna cada dia menos expressiva com o desenvolvimento de tecnologias de sensores, tornando-os mais baratos e possibilitando o seu uso em maior quantidade. Adicionalmente, diversas pesquisas recentes visam otimizar o posicionamento de sensores (Tan; Zhang, 2020) para maior eficiência no processo de detecção

Quadro 1 – Revisão de métodos de vibração de detecção de dano baseado em AI.

Referência	Abordagem	Estruturas testadas	Tipo de dano	Estratégias	Nível de SHM
Alves e Cury (2021)	DS	Viga, pórticos e ponte.	Furos, remoção de parafusos e rachaduras.	Características extraídas manualmente	I e II
Alves e Cury (2023)	DS	Viga, pórticos e pontes	Furos, rachaduras assentamento de fundação	Características extraídas manualmente e FS	I, II e III
Ghiasi, Ghaseini e Chan (2020)	ML	Pórticos, barras e ponte.	Não linearidades, adição de massa e retirada de intertravamentos.	FS e SVM	I
Finotti, Barbosa, Cury e Gentile (2017)	ML	Torre Gabbia (Itália)	Terremoto	SVM	I
Mariniello et al (2021)	ML	Pórtico, treliça e QUGC.	Retirada de intertravamento e afrouxamento de parafusos.	DT	I e II
Gui et al (2017)	ML	Pórtico.	Não-linearidades.	AR, SVM, GA, PSO	I e II
Zacharakis e Giagopoulos (2022)	ML	Chassi de veículo e barra de material composto.	Massa adicional e diminuição de rigidez.	PSO	I e II
Brusa et al. (2023)	ML	Máquinas rotatórias	Falhas no rolamento	SVM e kNN	I
Abdeljaber et al (2017)	DL	QUGS.	Afrouxamento de parafusos.	CNN	I e II

Luo <i>et al.</i> (2018)	DL	CNC	Falha precoce	Stacked AE, BPNN	I
Finotti, Barbosa, Cury e Pimentel (2021)	DL	Viga numérica e viaduto Várzea Nova em João Pessoa (Brasil)	Redução do módulo de elasticidade e reforço estrutural	SAE e SVM	I
Sony et al (2022)	DL	Ponte e QUGS.	Assentamento de fundação e afrouxamento de parafusos.	LSTM	I, II e III
Alazzawi e Wang (2022)	DL	Pórtico e Ponte.	Retirada de intertravamento e Cortes.	DRN	I, II e III
Chamangard (2022)	DL	QUGS e pontes.	Trincas e afrouxamento de parafusos.	CNN e TL	I
Luceli, Avci e Catbas (2023a) (2023b) (2023c)	DL	QUGS	Afrouxamento de parafusos.	CycleGAN	I
Amin <i>et al.</i> (2023)	DL	Turbinas eólicas	Falhas nas engrenagens	CNN	I

Quadro 1 – Revisão de métodos de vibração de detecção de dano baseado em AI.
Fonte: Elaboração própria.



Considerações finais

Esta pesquisa buscou apresentar estudos relevantes e recentes acerca das aplicações de AI no âmbito do SHM, utilizando unicamente dados de vibração para a detecção de danos estruturais. O presente trabalho se ateve a apresentar as diferenças entre abordagens de ML e DL, bem como as suas respectivas vantagens e limitações. Outrossim, destacou-se como os pesquisadores estão lidando atualmente com esses fatores adversos, o que elucida possibilidades de lacunas para trabalhos futuros. Note que em uma primeira abordagem ML, é necessário extrair características manualmente a partir dos sinais de aceleração, enquanto a segunda abordagem (DL) realiza a extração de maneira automática. Entende-se que diversos métodos implementados com AI foram bem-sucedidos quanto aos paradigmas de identificação da existência de dano, localização e/ou quantificação os quais se propuseram a solucionar.

No entanto, a dificuldade em coletar e rotular amostras válidas frequentemente leva a um pequeno conjunto de dados na identificação prática de danos, reduzindo o desempenho do modelo treinado. Para essa questão, foi relatada a implementação de modelos em elementos finitos, aprendizado por transferência e até redes neurais generativas. Além disso, modelos convencionais de aprendizado profundo muitas vezes carecem de um certo grau de interpretabilidade física. Tais desafios foram mencionados e discutidos como os pesquisadores têm lidado para eliminar ou mitigar essas limitações, por exemplo o uso de técnicas de IA explicáveis como LIME e SHAP.

Um aspecto comumente observado na literatura é o particionamento dos sinais de aceleração para gerar as amostras a serem treinadas pelo classificador com uma representatividade significativa. Outra similaridade encontrada é a recorrência das estruturas testadas. Isso se deve ao fato de serem estruturas do tipo *benchmark*, bem conhecidas e com dados disponíveis à comunidade científica. A seguir, são listadas as bases de dados mais comumente utilizadas em artigos dessa área e notadas pelos autores:

- IASC-ASCE (Johnson *et al.*, 2004);
- Viaduto Z24 (Maeck; De Roeck, 2003);
- Pórtico de três andares - Laboratório Nacional de Los Alamos (Figueiredo *et al.*, 2009);

- Ponte Tianjin Yonghe (Li *et al.*, 2012);
- QUGS (Abdeljaber *et al.*, 2017) ilustrado na Figura 5.

Assim, a utilização dessas bases possibilita a fácil comparação com outras metodologias de detecção, além de apresentar maior credibilidade e confiabilidade à pesquisa, pois já foram validadas anteriormente. Portanto, os autores recomendam, como boa prática, a utilização de alguma das estruturas listadas em trabalhos futuros que investiguem novas estratégias de SHM. Entretanto, novos dados de vibração experimentais de estruturas específicas são sempre necessários e relativamente difíceis de serem encontrados, como o caso de estruturas offshore, aeroespaciais, turbinas de geração de energia eólica, entre outras, o que revela uma ampla gama de setores da indústria ainda pouco explorados.

Agradecimentos

Os autores deste trabalho agradecem à UFJF (Universidade Federal de Juiz de Fora), à FAPEMIG (Fundação de Amparo à Pesquisa no Estado de Minas Gerais, projeto PPM-00001-18) e ao CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico, projetos 303982/2022-5 e 402533/2023-2) pelo auxílio financeiro.

Referências

- ABDELJABER, O.; AVCI, O.; KIRANYAZ, S.; GABBOUJ, M.; INMAN, D. J. Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks. *Journal of Sound and Vibration*, v. 388, p. 154-170, 2017. DOI:10.1016/j.jsv.2016.10.043
- ALAZZAWI, Osama; WANG, Dansheng. A novel structural damage identification method based on the acceleration responses under ambient vibration and an optimized deep residual algorithm. *Structural Health Monitoring*, v. 21, n. 6, p. 2587-2617, 2022. DOI:10.1177/14759217211065009
- ALVES, Victor; CURY, Alexandre. A fast and efficient feature extraction methodology for structural damage localization based on raw acceleration measurements. *Structural Control and Health Monitoring*, v. 28, n. 7, p. e2748, 2021. DOI:10.1002/stc.2748

ALVES, Victor; CURY, Alexandre. An automated vibration-based structural damage localization strategy using filter-type feature selection. *Mechanical Systems and Signal Processing*, v. 190, p. 110145, 2023. DOI:10.1016/j.ymssp.2023.110145

AMIN, A.; BIBO, A.; PANYAM, M.; TALLAPRAGADA, P. Wind Turbine Gearbox Fault Diagnosis Using Cyclostationary Analysis and Interpretable CNN. *Journal of Vibration Engineering & Technologies*, p. 1-11, 2023. DOI: 10.1007/s42417-023-00937-1

ARRIETA, Alejandro Barredo et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information fusion*, v. 58, p. 82-115, 2020. DOI: 10.1016/j.inffus.2019.12.012

AVCI, O.; ABDELJABER, O.; KIRANYAZ, S.; HUSSEIN, M.; GABBOUJ, M.; INMAN, D. J. A review of vibration-based damage detection in civil structures: From traditional methods to Machine Learning and Deep Learning applications. *Mechanical systems and signal processing*, v. 147, p. 107077, 2021. DOI:10.1016/j.ymssp.2020.107077

AZIMI, Mohsen; ESLAMLOU, Armin Dadras; PEKCAN, Gokhan. Data-driven structural health monitoring and damage detection through deep learning: State-of-the-art review. *Sensors*, v. 20, n. 10, p. 2778, 2020. DOI: 10.3390/s20102778

BRUSA, E.; CIBRARIO, L.; DELPRETE, C.; DI MAGGIO, L. G. Explainable AI for machine fault diagnosis: understanding features' contribution in machine learning models for industrial condition monitoring. *Applied Sciences*, v. 13, n. 4, p. 2038, 2023. DOI: 10.3390/app13042038

CHAMANGARD, M.; GHODRATI AMIRI, G.; DARVISHAN, E.; RASTIN, Z. Transfer Learning for CNN-Based Damage Detection in Civil Structures with Insufficient Data. *Shock and Vibration*, v. 2022, 2022. DOI: 10.1155/2022/3635116

CIREŞAN, D. C.; MEIER, U.; GAMBARDILLA, L. M.; SCHMIDHUBER, J. Deep, big, simple neural nets for handwritten digit recognition. *Neural computation*, v. 22, n. 12, p. 3207-3220, 2010. DOI: 10.1162/NECO_a_00052

CURY, Alexandre; CRÉMONA, Christian; DIDAY, Edwin. Application of symbolic data analysis for structural modification assessment. *Engineering Structures*, v. 32, n. 3, p. 762-775, 2010. DOI:10.1016/j.engstruct.2009.12.004

CURY, Alexandre; RIBEIRO, Diogo; UBERTINI, Filippo; TODD, Michael D. Structural health monitoring based on data science techniques. Springer, 2022. DOI:10.1007/978-3-030-81716-9

DANESHJOO, Z.; SHOKRIEH, M. M.; FAKOOR, M. A micromechanical model for prediction of mixed mode I/II delamination of laminated composites considering fiber bridging effects. *Theoretical and Applied Fracture Mechanics*, v. 94, p. 46-56, 2018. DOI: 10.1016/j.tafmec.2017.12.002

DONG, Chuan-Zhi; CATBAS, F. Necati. A review of computer vision-based structural health monitoring at local and global levels. *Structural Health Monitoring*, v. 20, n. 2, p. 692-743, 2021. DOI: 10.1177/1475921720935585

FINOTTI, R. P.; DE SOUZA BARBOSA, F.; CURY, A. A.; GENTILE, C. A novel natural frequency-based technique to detect structural changes using computational intelligence. *Procedia engineering*, v. 199, p. 3314-3319, 2017. DOI:10.1016/j.proeng.2017.09.438

- FINOTTI, R. P.; BARBOSA, F. D. S.; CURY, A. A.; PIMENTEL, R. L. Numerical and Experimental Evaluation of Structural Changes Using Sparse Auto-Encoders and SVM Applied to Dynamic Responses. *Applied Sciences*, v. 11, n. 24, p. 11965, 2021. DOI:10.3390/app112411965
- FIGUEIREDO, E.; PARK, G.; FIGUEIRAS, J.; FARRAR, C.; WORDEN, K. Structural health monitoring algorithm comparisons using standard data sets. Los Alamos National Lab. (LANL), Los Alamos, NM (United States), 2009. DOI: 10.2172/961604
- GHIASI, Ramin; GHASEMI, Mohammad Reza; CHAN, Tommy HT. Optimum feature selection for SHM of benchmark structures using efficient AI mechanism. *Smart Struct. Syst.*, v. 27, p. 623-640, 2021. DOI: 10.12989/sss.2021.27.4.623
- GUI, G.; PAN, H.; LIN, Z.; LI, Y.; YUAN, Z. Data-driven support vector machine with optimization techniques for structural health monitoring and damage detection. *KSCE Journal of Civil Engineering*, v. 21, p. 523-534, 2017. DOI: 10.1007/s12205-017-1518-5
- JOHNSON, E. A.; LAM, H. F.; KATAFYGIOTIS, L. S.; BECK, J. L. Phase I IASC-ASCE structural health monitoring benchmark problem using simulated data. *Journal of engineering mechanics*, v. 130, n. 1, p. 3-15, 2004. DOI: 10.1061/(ASCE)0733-9399(2004)130:1(3)
- Li, H.; Li, S.; Ou, J.; Li, H. Reliability assessment of cable-stayed bridges based on structural health monitoring techniques. *Structure and Infrastructure Engineering*, v. 8, n. 9, p. 829-845, 2012. DOI: 10.1080/15732479.2010.496856
- LULECI, Furkan; CATBAS, F. Necati; AVCI, Onur. Generative adversarial networks for labeled acceleration data augmentation for structural damage detection. *Journal of Civil Structural Health Monitoring*, v. 13, n. 1, p. 181-198, 2023a.
- LULECI, Furkan; AVCI, Onur; CATBAS, F. Necati. Improved undamaged-to-damaged acceleration response translation for Structural Health Monitoring. *Engineering Applications of Artificial Intelligence*, v. 122, p. 106146, 2023c. DOI: 10.1016/j.engappai.2023.106146
- LULECI, Furkan; CATBAS, F. Necati; AVCI, Onur. CycleGAN for undamaged-to-damaged domain translation for structural health monitoring and damage detection. *Mechanical Systems and Signal Processing*, v. 197, p. 110370, 2023b. DOI:10.1016/j.ymssp.2023.110370
- LUO, B; WANG, H.; LIU, H.; LI, B.; PENG, F. Early fault detection of machine tools based on deep learning and dynamic identification. *IEEE Transactions on Industrial Electronics*, v. 66, n. 1, p. 509-518, 2018. DOI: 10.1109/TIE.2018.2807414
- MARINIELLO, G.; PASTORE, T.; MENNA, C.; FESTA, P.; ASPRONE, D. Structural damage detection and localization using decision tree ensemble and vibration data. *Computer-Aided Civil and Infrastructure Engineering*, v. 36, n. 9, p. 1129-1149, 2021. DOI:10.1111/mice.12633
- MEIXEDO, A.; SANTOS, J.; RIBEIRO, D.; CALÇADA, R.; TODD, M. Damage detection in railway bridges using traffic-induced dynamic responses. *Engineering Structures*, v. 238, p. 112189, 2021. DOI:10.1016/j.engstruct.2021.112189
- MORALES, Fabricio A. O.; CURY, Alexandre A. Analysis of thermal and damage effects over structural modal parameters. *Structural engineering and mechanics: An international journal*, v. 65, n. 1, p. 43-51, 2018. DOI:10.12989/sem.2018.65.1.043

- MAECK, Johan; DE ROECK, Guido. *Damage assessment using vibration analysis on the Z24-bridge. Mechanical Systems and Signal Processing*, v. 17, n. 1, p. 133-142, 2003. DOI: 10.1006/mssp.2002.1550
- MOUGHTY, John J.; CASAS, Joan R. *A state of the art review of modal-based damage detection in bridges: Development, challenges, and solutions. Applied Sciences*, v. 7, n. 5, p. 510, 2017. DOI: 10.3390/app7050510
- RYTTER, A. *Vibrational based inspection of civil engineering structures*. 1993. Tese de doutorado – Denmark: Department of Building Technology and Structural, Aalborg University, Aalborg, 1993.
- SONY, S.; GAMAGE, S.; SADHU, A.; SAMARABANDU, J. *Vibration-based multiclass damage detection and localization using long short-term memory networks. Structures*, v. 35, p. 436-451, 2022. DOI: 10.1016/j.istruc.2021.10.088
- TAN, Yi; ZHANG, Limao. *Computational methodologies for optimal sensor placement in structural health monitoring: A review. Structural Health Monitoring*, v. 19, n. 4, p. 1287-1308, 2020. DOI: 10.1177/1475921719877579
- TIAN, W.; CHENG, X.; LIU, Q.; YU, C.; GAO, F.; CHI, Y. *Meso-structure segmentation of concrete CT image based on mask and regional convolution neural network. Materials & Design*, v. 208, p. 109919, 2021. DOI:10.1016/j.matdes.2021.109919
- WANG, Teng; LU, Guoliang; YAN, Peng. *A novel statistical time-frequency analysis for rotating machine condition monitoring. IEEE Transactions on Industrial Electronics*, v. 67, n. 1, p. 531-541, 2019. DOI: 10.1109/TIE.2019.2896109
- YUAN, Fuh-Gwo (Ed.). *Structural health monitoring (SHM) in aerospace structures*. Woodhead Publishing, 2016. DOI: 10.1016/C2014-0-00994-X
- ZACHARAKIS, Ilias; GIAGOPOULOS, Dimitrios. *Vibration-Based Damage Detection Using Finite Element Modeling and the Metaheuristic Particle Swarm Optimization Algorithm. Sensors*, v. 22, n. 14, p. 5079, 2022. DOI: 10.3390/s22145079
- ZHANG, Yixiao; LEI, Ying. *Data anomaly detection of bridge structures using convolutional neural network based on structural vibration signals. Symmetry*, v. 13, n. 7, p. 1186, 2021. DOI: 10.3390/sym13071186





 Giselle Beiguelman, frames do vídeo generativo da série Flora Rebellis, versão 2024.
Imagens geradas com Inteligência Artificial (Style Gans)

 Acesse o vídeo clicando na imagem





PAY ATTENTION: THE HIGH-SPEED EVOLUTION OF NLP, AND WHERE IT HITS A WALL

DOI: <https://doi.org/10.35699/2965-6931.2023.47510>

FÁBIO COZMAN*; HUGO NERI**

ABSTRACT: This paper analyzes the evolution of attention-based models in Natural Language Processing (NLP) with an informal tone, starting from 2003 and culminating in the transformer architectures we know since 2017. We explain how transformers have managed to solve significant benchmarks for commonsense reasoning in Artificial Intelligence due to their pre-training. Further, we investigate the parallel between the concept of 'gist' in human language understanding, as proposed by Roger Schank, and the 'embeddings' now employed in machine learning. Towards the end of the paper, we discuss a well-known problem with these models, the so-called "hallucinations." This phenomenon highlights the models' struggle to discern fact from fiction, necessitating further research to mitigate its impact. We frame this issue in the context of David Lewis's work, arguing that it represents a fundamental challenge for language models.

KEYWORDS: language models; deep learning; natural language processing; artificial intelligence.

Preste Atenção: A veloz evolução do processamento de linguagem natural e onde ela empaca

RESUMO: Este artigo analisa a evolução dos modelos baseados em atenção no Processamento de Linguagem Natural (PLN) em um tom informal, começando em 2003 e culminando nas arquiteturas de "transformers" que conhecemos desde 2017. Explicamos como os "transformers" conseguiram resolver o importante "benchmark" para o raciocínio de senso comum em Inteligência Artificial devido ao seu pré-treinamento. Além disso, investigamos o paralelo entre o conceito de 'gist' ("o que realmente importa") na compreensão da linguagem humana, conforme proposto por Roger Schank, um veterano do PLN, e os "embeddings" agora empregados na aprendizagem de máquina. No final do artigo, discutimos um problema bem conhecido com esses modelos, as chamadas "alucinações". Este fenômeno destaca a luta dos modelos para discernir o fato de ficção, necessitando de mais pesquisas para mitigar seu impacto. Enquadramos essa questão no contexto do trabalho de David Lewis, argumentando que representa um desafio fundamental para os modelos de linguagem.

PALAVRAS-CHAVE: modelos de linguagem; aprendizado profundo; processamento de linguagem natural; inteligência artificial.

* Universidade de São Paulo / ** Universidade de São Paulo

A new predator in the nlp landscape

In the Darwinian landscape of machine learning, the advent of the attention mechanism led to the emergence of a new species of computational model for language processing. Similarly to the most adaptable organisms in the natural world, these models have evolved at first slowly, then dramatically, reshaping the landscape of natural language processing. The genesis of this novel species can be traced back to the concept of dynamically adjusting focus within sequences, mirroring how the human brain selectively attends to relevant *stimuli*. This idea appears in the groundbreaking paper “Neural Machine Translation by Jointly Learning to Align and Translate” by Bahdanau et al (2016).

Predating this, Bengio et al. (2003) laid the bedrock for the attention mechanism. They conceived a model that learned a distributed representation for each word and the probability function for word sequences simultaneously, employing a methodology evocative of how attention weights words based on their contextual relevance. Intuitively: they developed a method that gives each word a kind of ‘tag’ or ‘score’ that reflects its relationships with other words (the embedding of the word). At the same time, the method calculates how often certain words follow each other in a sequence. The resulting model does not truly understand the words or their meanings, but it can predict which words are likely to come next in a series (e.g., a text), perhaps the same way humans themselves predict language. Their exploration of vector space representations for words presaged the concept of word embeddings, a cornerstone in attention-based models. While their model did not fully implement the concept of dynamic input weighting based on the current context, the focus on a ‘short list’ of high-probability words hinted at an early, rudimentary form of attention.

In the evolution of NLP the advent of word embeddings has been a major milestone. You can think of word embeddings as a skill that these machine learning models acquired. It is similar to a tracking system, where each word gets a unique numeric code, similar to animal tracks in the wild. Just like tracks can help predict

the movements of an animal, these numeric codes help the models predict words that frequently appear together or are contextually related.

Imagine these numeric codes as having multiple dimensions, similarly to how animal tracks provide information about the size, species, or speed of the animal. In the case of word embeddings, each 'dimension' captures a different aspect of how a word interacts with other words in a language. The more dimensions we can track, the more information we have about the word's behavior in its natural habitat of language. This multi-faceted tracking system equips machine learning models with a powerful tool to navigate the complex world of natural language processing (Bengio et al., 2003; Mikolov et al., 2013a; Mikolov et al., 2013b; Pennington; Socher; Manning, 2014; Levy; Goldberg, 2014).

The attention mechanism, taking root from these foundational concepts, matured to surpass the constraints of previous NLP techniques. The mechanism can master long sequences, refine its understanding of context, and excel in parallel token processing. Equipped with such skills, the attention mechanism has tackled a diverse array of natural language problems that previous models struggled with.

A significant leap in the evolution of the attention mechanism came with the 2017 publication of Vaswani et al.'s paper, "Attention is All You Need" (2017). This paper began a radical shift from traditional architectures by positing that attention could form the bedrock of neural network architectures, thereby eliminating the need for recurrent and convolutional layers. This evolutionary leap instigated a dramatic reshaping, comparable to the emergence of a new apex predator that alters the balance of an ecological community. The Transformer model, a novelty introduced in that revolutionary paper, harnessed the power of the self-attention mechanism. This approach allowed the model to weigh and incorporate all tokens in the input sequence, irrespective of their positions, thus enabling it to establish dependencies between tokens far apart in the sequence. The Transformer model also introduced a novel positional encoding scheme to account for word order, a crucial element in language understanding.

The Transformer model's scalability, coupled with its unparalleled capacity to model intricate patterns in data, paved the way for the development of larger and more powerful models. This led to a new generation of language models such as GPT (Radford et al., 2019) and Bert (Devlin et al., 2018), and Roberta (Liu et al., 2019) whi-

1 Binary Decision

Trees are a form of machine learning model that makes decisions based on a series of yes/no questions about the data. Each node in the tree represents a condition or question about the data, and each branch represents the outcome of the question, leading to different paths down the tree. The final decision or prediction is made at the leaf nodes.

2 Hidden Markov

Models are statistical models where you observe a series of events but do not know the exact sequence of underlying states that led to those events. It uses probabilities to predict the sequence of hidden states.

3 The LSTM is a type of recurrent neural network that can remember or forget information over time.

ch have since achieved unprecedented performance across a broad spectrum of NLP tasks, from translation to text generation. Complex tasks such as machine translation, text summarization, and sentiment analysis succumbed to their prowess, and even longstanding challenges like the Winograd challenge, paraphrasing challenges, even arguments classification, have been surmounted by this apex predator.

The ‘prey’: natural language challenges

The landscape of natural language processing is teeming with diverse ‘prey,’ or challenges, that continually test the prowess of the corresponding AI ‘predators.’ The Transformer model, the current apex predator, has shown remarkable adaptability and effectiveness, conquering various language challenges once thought insurmountable. For instance, the Transformer’s expansive ‘hunting ground’ includes question answering, a task demanding a deep understanding of context and accurate information retrieval. On that terrain, the Transformer has demonstrated exceptional skill, exceeding previous performance standards on benchmarks such as the Squad dataset (Rajpurkar et al., 2016), a reading comprehension dataset consisting of questions posed by crowdworkers on a set of Wikipedia articles.

In the task of named entity recognition, earlier hunters such as Binary Decision Trees¹ (Bennett et al., 1997) and Hidden Markov Models² (Bikel et al., 1999) have been superseded by Transformers. Even more recent ones like Long Short-Term Memory (LSTM³) (Lample et al., 2016), which despite being state-of-the-art for a time, have their relevance threatened. Transformers have left their mark by identifying and categorizing key information in text, such as the names of people, organizations, locations, and more. This task, analogous to tracking and classifying different types of prey, speaks to the Transformer’s adeptness in understanding relationships between different words in a sentence. Transformers have also succeeded in the intricate territory of Natural Language Inference (NLI) with the Stanford Natural Language Inference (SNLI) corpus (Bowman et al., 2015). Much like navigating a complex terrain to capture elusive prey, NLI requires a delicate balance of language understanding and logical reasoning. Transformers, with their powerful mechanism for capturing

long-range dependencies, have set new performance records on the Multinli dataset (Williams et al., 2018).

Transformers have also demonstrated their versatility by excelling in text classification, effectively categorizing text into predefined categories with remarkable accuracy.

Even the elusive Winograd family of problems, from WS-273 Davies (2016, 2017), Levesque (2011, 2013, 2017), and Levesque et al. (2012) to WinoGrande (Sakaguchi et al., 2019), possibly the wildest of prey, has not remained unscathed by the Transformer. These problems of pronoun resolution, usually straightforward for humans but complex for machines, have been in essence solved by Transformers, thus illustrating their adeptness at nuanced and common-sense understanding.

As victors in the vast landscape of natural language processing, Transformers have vanquished many once-daunting challenges. They have shown remarkable dexterity in understanding human language, not by mimicking human cognitive processes but by displaying an uncanny knack for detecting patterns and relationships within colossal amounts of text data. This has resulted in pragmatic effects that outperform many traditional approaches and often approximate human-level performance.

Their victories are not confined to small battles; they have managed to master complex tasks like question answering, named entity recognition, and natural language inference. These achievements highlight the Transformer current dominance, but also pose new questions, akin to the mysteries that arise after every significant conquest.

The remaining natural language challenges are survivors indeed, but they are not merely remnants of a vanquished enemy. They represent new, exciting frontiers, challenges that expose the inherent complexity of human language and cognition, and the intricate nature of our world. These are not problems to be dismissed or understated, but intriguing puzzles that continue to captivate and challenge us. Despite their elusiveness, these problems do not pose insurmountable hurdles. They are akin to the hidden but not unreachable game in a savanna. Our engineering ingenuity, constantly evolving methodologies, and creative problem-solving can help us track these elusive targets.

The Transformers' victories, while astonishing, have also shaped the landsca-

pe of natural language processing in unexpected ways. They have allowed us to look into more abstract concepts and ideas, pushing the boundaries of what we once thought was achievable.

In the next sections of this paper, we will delve deeper into details behind the Transformers' success and explore the intriguing issue of hallucinations in language models. The hunt continues, but now it is about tracking down the elusive, complex beasts of the NLP savanna. These remaining challenges symbolize not just the limits of our current understanding, but also the fascinating possibilities for future breakthroughs.

Transformers (pre-) training is commonsense training

Transformers have succeeded in cracking several challenging problems connected to commonsense reasoning. The journey towards understanding this success requires a foray into the roots of Bert and Roberta's training process, masked language modeling (Neri & Cozman, 2023).

BERT, focusing on bidirectional language learning, masks a certain portion of words during its training phase. This adaptation mirrors a unique survival tactic in the linguistic wilderness — the cloze test. First introduced by Wilson Taylor in 1953, this test assesses communication effectiveness by asking individuals to fill in the blanks, completing the sentence much like closing gaps in a lineage. The parallel between Bert's masked language modeling and the cloze test is clear and striking.

Linguistic commonsense, the skill needed to complete a cloze unit, equips an individual with the ability to understand and complete mutilated sentences. This capability, which interacts with various enabling and obstructing elements, evolves into our comprehensive understanding of linguistic commonsense.

Given this, Bert's training can be viewed as a large-scale implementation of the cloze procedure, fostering the development of linguistic commonsense within a machine.

We should note that Winograd schemas (Davies, 2016; Davies et al., 2017) mi-

mic cloze units in their environment. In fact, the Winograd Schema Challenge (WSC) demands this evolved skill, as it requires deducing anticipated words from their linguistic context. It is clear that Bert's evolutionary approach is well-suited for the challenges posed by the WSC.

Further into the linguistic wilderness, the Recognizing Textual Entailment (RTE) task, akin to the WSC, tests for Textual Entailment (TE) within sentences (Dagan, 2009). Despite RTE's reliance on inferring causal relationships, Levesque suggested the WSC as a solution, introducing a mutation of double ambiguity, resulting in a cloze test-like environment (Levesque, 2011).

In this evolution-inspired narrative, the WinoGrande corpus and Roberta play key roles (LIU et al., 2019). Roberta, a more advanced version of BERT, has demonstrated significant leaps in performance within the WSC. This is attributed to superior pretraining involving larger corpora and more effective hyperparameters. In addition to the data sources used to train Bert, Roberta also incorporated the English portion of the CC-NEWS dataset, OpenWebText, and Stories from CommonCrawl.

The fine-tuning of Roberta using the Definite Pronoun Resolution Dataset (DPR) led to a leap in its fitness, reaching an impressive 83.1% accuracy in the WSC. This adaptation was further honed with the integration of WinoGrande, enhancing accuracy to an even more impressive 90.1%. Similar adaptive improvements were observed in the Pronoun Disambiguation Problems (PDP) dataset. Unexpectedly, the solution of the Choice of Plausible Alternatives (COPA) saw a substantial leap. Here, the expansive WinoGrande corpus provided a diverse environment conducive for training, with RoBERTa's performance soaring with the introduction of over 40k WinoGrad schemas.

While Roberta's progressive advancements and the vast WinoGrande environment are significant, some critics argue that these strides might be the result of evaluation anomalies and artifacts, rather than a genuine evolution in LLMs (Gururangan et al., 2018; Sakaguchi et al., 2019b). Regardless, the pivotal role of pretrained LLMs in enhancing word sense disambiguation is unquestionable, as demonstrated by the marked performance leap across different iterations of the WinoGrande project (Sakaguchi et al., 2019a).

Why are we confident that the masked language task during pretraining was

the true driving force behind the impressive 90.1% performance observed in Sakaguchi et al. (2019b)? It is because an exceptionally large model such as GPT-3 did not achieve the anticipated stellar performance on the WSC: our tests resulted in a modest 67% accuracy (Neri & Cozman, 2023). Performance improved for the even larger GPT-4, yielding an accuracy of 85.2%. While commendable, this performance pales when compared to the models' sizes: GPT-3.5 has 175B parameters, and GPT-4 is even larger, though its exact size remains unknown (Openai, 2023). Conversely, Bert and Roberta have only 340 million parameters each. Indeed, size matters, but skill plays a far more crucial role. This represents a specialization within the current transformers landscape, with the Bert family rooted in pretraining that resembles a human reading task, while the GPT family simulates an oral task, where the next token prediction is paramount (Neri & Cozman, 2023).

Transformers capture, in their own way, the gist of language

The notable cognitive scientist Roger Schank, in collaboration with Robert P. Abelson, substantially advanced the early field of Natural Language Processing (NLP) with their pioneering publication, "Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures" (1977). Their work established a foundational understanding of language processing and generation, both by humans and machines. Schank's distinctive view, centered on the intersection of AI and human cognition, brought a fresh perspective, even though it was not rooted in conventional machine learning or connectionist theories. Their perspective highlighted the necessity of cognitive frameworks in AI, prefiguring many subsequent discussions in NLP.

In the 1990s, Schank shifted his focus to education, thereby further augmenting his grasp of learning processes. The resonance and relevance of his initial work on AI persist in today's AI dialogue. His book from the '90s, "Tell Me a Story: Narrative and Intelligence," offers pertinent insights into our discourse on evolving more human-like AI. Schank postulates in that book that the essence of human intelligence and memory formation lies in storytelling and narratives, which suggests a unique

approach to training AI systems to comprehend and create narratives, hence augmenting their ‘human-like’ attributes. This is especially pertinent to this paper as we examine how AI models deduce the ‘gist’ or underlying narrative patterns, akin to human cognition.

Schank proposes that storytelling forms the bedrock of human intelligence, asserting that narratives help us organize our memories, grasp our surroundings, and shape our life experiences. This extends beyond mere communication; narratives become the prism through which we interpret our world.

This perspective carries significant implications within Artificial Intelligence. It indicates that advancing towards a more human-like AI might depend on equipping machines with the capacity to understand and weave narratives. This is how large language models in fact function; they generate responses based on patterns and structures they have discerned from a massive spectrum of narratives in their training data.

Schank’s theory emphasizes the concept of the “gist,” the fundamental essence or core meaning of a story. According to Schank, we process and remember stories by distilling these main themes or schemas, abstract structures representing our knowledge and expectations. When we listen to a story, we do not remember every detail but retain the gist — the basic structure and the key elements that lend meaning to the story. To an extent, AI language models replicate this process.

This idea of a “gist” in Schank’s theory exhibits parallels to “embeddings” in machine learning. Embeddings are mathematical constructs where similar entities are grouped closely. In the context of natural language processing, words or phrases with similar meanings are represented as vectors in a multidimensional space, situated near each other. Just as the “gist” encapsulates a story’s essential meaning, these embeddings distill the fundamental connotation of a word or phrase.

The ability of embeddings to capture semantic similarities is crucial in recognizing synonyms across different dialects or jargons. For example, a sophisticated language model can comprehend that ‘car’ and ‘automobile’ or ‘solicitor’ and ‘lawyer’ carry similar meanings, even when used in varying contexts. This ability is critical in applications like information retrieval or machine translation, as it allows accurate mapping of words and phrases across languages.

Nonetheless, embeddings, like all tools, have their constraints. Their effica-

cy hinges on the quality and diversity of the training data. If certain words, phrases, or concepts are underrepresented, the embeddings might not accurately encapsulate their meanings or subtleties. This underlines the significance of a comprehensive, diverse, and high-quality dataset for training more nuanced and effective language models.

Llms can explore the unreal

Large language models such as GPT-4 have the ability to engage with hypothetical scenarios and counterfactuals, offering a fascinating glimpse into the world of “what ifs”. This ability, integral to human cognition, allows for exploration of uncharted territory, evaluation of different possibilities, and anticipation of future scenarios. Advanced language models like GPT-4 can create responses to these hypothetical situations based on patterns learned from vast text data. However, their responses lack the genuine emotional understanding or empathy that underlies human responses.

The concept of Counterpart Theory by David Lewis provides a useful perspective here. In that theory, entities in different possible worlds can be ‘counterparts’ with similar characteristics but not identical (LEWIS, 1968). The responses generated by AI models to hypothetical scenarios can be considered ‘counterparts’ to human responses — they imitate human behavior but remain fundamentally different due to the absence of subjective experiences, emotions, and ethical comprehension.

Lewis’s notion of truth within fictional narratives is another significant aspect. He suggests that fictional stories create their own ‘possible worlds’, and statements about these worlds can be considered true within the confines of those worlds (Lewis, 1978). Language models, using the patterns, relationships, and rules they have learned, can generate narratives that are coherent within the possible world of a given fiction.

The Principle of Minimal Departure proposed by Lewis states that when a story leaves certain details unspecified, we fill these gaps based on our real world, assuming they do not contradict the narrative (Lewis, 1968). Language models often adopt a similar approach when generating text, filling in the blanks with patterns and facts learned from their training data. The fact that truth in fiction is context-dependent,

according to Lewis, presents a significant challenge and opportunity for large language models. While they are capable of producing varied responses depending on context, their sensitivity to context relies more on pattern recognition than on genuine understanding.

Future developments in this area may answer intriguing questions like: Can large language models generate more sophisticated responses to counterfactuals? Will they better leverage context to navigate the nuances of truth within fiction? As they continue to evolve, will they further refine their mimicry of human-like understanding, or will they develop a unique way of interpreting these issues?

Parallel universes

David Lewis's philosophy is based on the concept of possible worlds, which represent comprehensive ways things could have been. This concept resonates with the universe of AI, especially language models like GPT-4, which live in a world of linguistic data devoid of physical matter. Their 'possible world' is constructed from patterns, statistical associations, and symbolic relationships encoded in the training data.

Lewis's philosophy espouses modal realism about possible worlds, considering each of these worlds as real as our actual world. This parallels with the 'world' of a language model which is constituted by linguistic data and relationships (Lewis, 1986). From Lewis's perspective, a plethora of possible worlds exist, and our actual world is merely one of these realities. This is similar to the functioning of language models like GPT-4, wherein each unique combination of input parameters and model states can lead to the generation of a different 'reality' or output.

Lewis further posits that these possible worlds exist independently and do not causally interact. Despite this isolation, we can conceptualize and communicate about these alternate worlds, enabling the creation of fiction. AI language models function in a similar manner, treating each input-output pair separately, with no causal interaction between different pairs. However, they can produce a variety of contextually appropriate responses, crafting narratives much like a fictional story. In the realm of AI, Lewis's counterpart theory can be related to the fact that each unique output from a model for a given input can be viewed as 'counterparts' — not identical but similar representations.

Hallucinations in language models: a perennial risk? transformers cannot distinguish between truth and fiction, actual and possible worlds

Hallucination in the context of LLMs such as GPT-4 refers to the generation of facts that are not related to a given input nor directly derived from the training data. This phenomenon typically surfaces when the model is tasked with filling in gaps or extrapolating on ambiguous inputs. As it strives to generate human-like text, it draws on its extensive training to make educated guesses, sometimes leading to outputs that, while seemingly plausible, are not necessarily grounded in reality.

A pivotal issue in this process lies in the LLM's inability to discern fiction from fact. These models are trained on a vast corpus of text data that comprises a blend of factual and fictional content. Absent an inherent understanding of the difference, they may inadvertently 'hallucinate' elements from their training data, weaving fictional elements into contexts that call for strictly factual information, and vice versa.

For instance, when asked about a historical event, an LLM might generate a response that blends factual data with fictional elements derived from a novel set in the same era. Though the generated text may appear coherent and contextually appropriate, it may misrepresent the actual event, thus generating a 'hallucination.'

The potential risks associated with such hallucinations are significant. In critical applications such as medical advice or legal counsel, hallucinations may lead to erroneous decisions with grave consequences. For example, an LLM trained on a mixture of real medical textbooks and science fiction novels might 'hallucinate' while providing medical advice, yielding recommendations that are scientifically unfounded and potentially harmful. In the same vein, a LLM can be tricked to do not distinguish a fictional conversation to a real one, allowing it to produce harmful speech.

So, how do we mitigate these hallucinations? Strategies include refining the quality and diversity of training data, implementing fact-checking mechanisms during the output generation process, and exploring techniques to limit the LLM's propensity to generate text beyond its training data. However, completely eliminating hallucinations remains a formidable (if not impossible) challenge, primarily due to the inherent

design and training methods of these models.

Hallucination, while a challenge, could also be perceived as a perennial trait of language models like GPT-4, leading to an intriguing paradox. Their strength — generating human-like text — can also be their Achilles' heel when the boundaries blur between fiction and reality. As we continue to navigate this complex landscape, these 'hallucinations' stand not just as a risk to be managed, but also as a compelling puzzle driving us towards the development of more sophisticated models. Models that can better distinguish fiction from reality, effectively reducing hallucinations, all the while refining our understanding of the vast and nuanced terrain of artificial intelligence.

Final remarks

Large Language Models (LLMs) have reached unprecedented heights in language-related tasks, often exhibiting capabilities that rival, or even surpass, human proficiency. The breadth of their linguistic knowledge and their ability to adapt to various contexts allow them to not only comprehend and generate language but to do so in a way that encapsulates the essence, or the “gist,” of the discourse at hand.

LLMs leverage ‘embeddings’, a powerful representation of words or phrases as vectors in a high-dimensional space. These embeddings are trained to capture the semantic essence of words, facilitating translation between languages by mapping similar concepts together. However, LLMs push this capability even further by translating not just across different languages, but also between audiences. This means they can take a concept expressed in specialized jargon and translate it into plain language, making complex topics accessible to a wider audience.

However, the prowess of LLMs is rooted solely in the realm of language. They exist in a world of linguistic data, devoid of physical matter. This inherent characteristic shapes their limitations as well as their strengths. A significant limitation that comes to the fore is the phenomenon of ‘hallucination.’ In the context of machine learning, hallucination refers to the generation of details that are not present in the input or training data. When an LLM faces a linguistic situation with missing or ambiguous

information, it often ‘fills in the gaps,’ which may lead to outputs that are unrelated to the original context or not based in reality. This makes the task of mitigating such undesired outputs an external endeavor, not inherent to the model itself.

The enormous size of LLMs, while contributing to their extraordinary capabilities, also poses challenges. These models require vast computational resources for training and operation, making them difficult to manage and potentially raising issues related to efficiency and environmental sustainability. Moreover, the complexity of these models can make understanding and predicting their behavior increasingly challenging. However, the discussion around these concerns is expansive and deserves a dedicated exploration, potentially opening new avenues of research and innovation.

References

- BAHDANAU, D.; CHO, K.; BENGIO, Y. *Neural Machine Translation by Jointly Learning to Align and Translate*. 2016. Disponível em: <https://arxiv.org/abs/1409.0473>.
- BENGIO, Y. et al. *A neural probabilistic language model*. *Journal of Machine Learning Research*, v. 3, n. Feb, p. 1137-1155, 2003.
- BENGIO, Y.; DUCHARME, R.; VINCENT, P.; JANVIN, C. *A neural probabilistic language model*. *Journal of Machine Learning Research*, v. 3, p. 1137-1155, 2003.
- BENNETT, S.W.; AONE, C.; LOVELL, C. *Learning to tag multilingual texts through observation*. In: *SECOND CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING*, 2., 1997, Providence. *Proceedings*. Providence: Morgan Kaufmann Publishers, Inc, 1997. p. 109-116.
- BOWMAN, S. R.; ANGELI, G.; POTTS, C.; MANNING, C. D. *A large annotated corpus for learning natural language inference*. 2015. Available at: <http://arxiv.org/abs/1508.05326>.
- BROWN, T. B.; et al. *Language models are few-shot learners*. *ArXiv:2005.14165*, 2020.
- DAGAN, Ido. *Recognizing textual entailment: Rational, evaluation and approaches*. *Natural Language Engineering*, v. 15, n. 4, p. i-xvii, 2009.
- DAVIES, Ernest. *Winograd schemas and machine translation*. 2016. Disponível em: <https://arxiv.org/>

abs/1608.01884.

DAVIES, Ernest; MORGENSTERN, Leora; ORTIZ, Charles L. *The first Winograd Schema Challenge at IJCAI-16*. *AI Magazine*, v. 38, n. 3, p. 97-98, 2017.

DEVLIN, J.; CHANG, M. W.; LEE, K.; TOUTANOVA, K. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. 2018. Available at: <https://arxiv.org/abs/1810.04805>.

GURURANGAN, Suchin; et. al. *Annotation Artifacts in Natural Language Inference Data*. 2018. Available at: <https://arxiv.org/abs/1803.02324>.

LEVESQUE, Hector. *The Winograd Schema Challenge*. In: *AAAI SPRING SYMPOSIUM*, 2011, Palo Alto. *Anais*. Palo Alto: AAAI, 2011.

LEVESQUE, Hector. *On our best behaviour*. In: *INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE*, 2013, Beijing. *Anais*. Beijing: IJCAI, 2013.

LEVESQUE, Hector. *Common Sense, the Turing Test, and the Quest for Real AI*. Cambridge, Massachusetts: The MIT Press, 2017.

LEVESQUE, Hector; DAVIES, Ernest; MORGENSTERN, Leora. *The Winograd Schema Challenge*. In: *PRINCIPLES OF KNOWLEDGE REPRESENTATION AND REASONING*, 2012, Rome. *Proceedings*. Rome: KR, 2012.

LEVY, O.; GOLDBERG, Y. *Neural word embedding as implicit matrix factorization*. In: *ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS*. *Anais*. 2014. p. 2177-2185.

LEWIS, D. *Counterpart theory and quantified modal logic*. *Journal of Philosophy*, v. 65, n. 5, p. 113-126, 1968.

LEWIS, D. *On the Plurality of Worlds*. Oxford: Blackwell, 1986.

LEWIS, D. *Truth in Fiction*. *American Philosophical Quarterly*, v. 15, n. 1, p. 37-46, 1978.

LIU, Y. et al. *RoBERTa: A Robustly Optimized BERT Pretraining Approach*. 2019. Available at: <https://arxiv.org/abs/1907.11692>.

MIKOLOV, T. et al. *Distributed representations of words and phrases and their compositionality*. In: *ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS*, *Anais*. 2013. p. 3111-3119.

MIKOLOV, T. et al. *Efficient estimation of word representations in vector space*. 2013. Disponível em: <https://arxiv.org/abs/1301.3781>.

OPEN AI. *GPT4 Technical Report*. *ArXiv*:2303.08774, 2023.

PENNINGTON, J.; SOCHER, R.; MANNING, C. *Glove: Global vectors for word representation*. In: *CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING*, 2014. *Proceedings*. 2014. p. 1532-1543.

RADFORD, A.; WU, J.; CHILD, R.; LUAN, D.; AMODEI, D.; SUTSKEVER, I. *Language models are unsupervised multitask learners*. *OpenAI Blog*, v. 1, n. 8, 2019.

RADFORD, A.; WU, J.; CHILD, R.; LUAN, D.; AMODEI, D.; SUTSKEVER, I. *Language Models are Unsupervised Multitask Learners*. *OpenAI Blog*, v. 1, n. 8, 2019. Available at: https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf.

RAJPURKAR, Pranav; ZHANG, Jian; LOPYREV, Konstantin; LIANG, Percy. *SQuAD: 100,000+ Questions for Machine Comprehension of Text*. In: *CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING*, 2016, Austin, Texas. *Anais*. Austin, Texas: Association for Computational Linguistics, 2016. p. 2383-2392.

RYAN, M. L. *Possible worlds, artificial intelligence, and narrative theory*. Bloomington: Indiana University Press, 1991.

SAKAGUCHI, K.; LE BRAS, R.; BHAGAVATULA, C.; CHOI, Y. *Winogrande: An adversarial Winograd Schema Challenge at scale*. In: *AAAI-20 TECHNICAL TRACKS* 5, 34, 2019. p. 05.

SCHANK, R.; ABELSON, R.P. *Scripts, plans, goals and understanding: An inquiry into human knowledge structures*. New Jersey: Erlbaum, 1977.

SCHANK, R.C. *Tell Me a Story: A New Look at Real and Artificial Memory*. 1st ed. New York: Atheneum, 1990.

TAYLOR, W. "Cloze procedure": A new tool for measuring readability. *Journalism Quarterly*, v. 30, n. 4, p. 415-433, fall, 1953.

VASWANI, A. et al. *Attention is all you need*. In: *ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS*. *Anais*, 2017. p. 5998-6008.

WILLIAMS, A.; NANGIA, N.; BOWMAN, S. *A Broad-Coverage Challenge Corpus for Sentence Understanding through Inference*. In: *CONFERENCE OF THE NORTH AMERICAN CHAPTER OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS: HUMAN LANGUAGE TECHNOLOGIES*, 2018, New Orleans. *Anais*. New Orleans: Association for Computational Linguistics, 2018. p. 1112-1122.





■ Giselle Beiguelman, frames do vídeo generativo da série Flora Rebellis, versão 2024.
Imagens geradas com Inteligência Artificial (Style Gans)

■ Acesse o vídeo clicando na imagem





HEURISTIC PROPOSAL FOR GENERATION OF VALID SOLUTIONS TO A FOREST PRODUCTION PLANNING PROBLEM

DOI: <https://doi.org/10.35699/2965-6931.2023.46940>

ISABELLA ALKMIM BARRAL*; LUCIANA BALIEIRO COSME***; ALLYSSON STEVE MOTA LACERDA***

ABSTRACT: In the forestry sector, efficient planning of activities is considered essential for companies to achieve their goals and remain competitive in the market. To assist in this process, this work proposes the development and evaluation of a heuristic for the generation of valid solutions for a forest production planning problem involving 120 stands and 81 management alternatives per field, totaling 9720 decision variables. In the tests carried out, such proposed heuristic presented promising results when compared to the literature, indicating that it is capable of providing good quality solutions with low computational cost. Results also indicate that the inclusion of solutions obtained by using the proposed heuristic in the set of initial individuals/antibodies of Genetic Algorithm and Clonal Selection Algorithm can generate a significant improvement in the quality of final solutions and, consequently, in the performance of such meta-heuristics.

KEYWORDS: forestry planning; heuristics; valid solution; metaheuristics.

Proposta de heurística para a geração de soluções válidas para um problema de planejamento da produção florestal

RESUMO: No setor florestal, o planejamento eficiente das atividades é considerado fundamental para que as empresas atinjam seus objetivos e permaneçam competitivas no mercado. Para auxiliar nesse processo, o presente trabalho propõe o desenvolvimento e avaliação de uma heurística para a geração de soluções válidas para um problema de planejamento da produção florestal envolvendo 120 talhões e 81 alternativas de manejo por talhão, totalizando 9720 variáveis de decisão. Nos testes realizados a heurística apresentou resultados promissores quando comparada com a literatura, indicando que é capaz de fornecer soluções de boa qualidade com um baixo custo computacional. Os resultados também apontam que a inclusão de soluções obtidas pela heurística proposta no conjunto de indivíduos/anticorpos iniciais do Algoritmo Genético e Algoritmo de Seleção Clonal pode gerar uma melhora significativa no desempenho dessas meta-heurísticas. palavras-chave: modelos de linguagem; aprendizado profundo; processamento de linguagem natural; inteligência artificial.

PALAVRAS-CHAVE:: planejamento florestal; heurística; solução válida; meta-heurísticas.

* Universidade Estadual de Montes Claros/ ** Instituto Federal do Norte de Minas Gerais/ *** Universidade Estadual de Montes Claros

Introduction

The Brazilian forestry sector, more specifically that of planted forests, plays a fundamental role in the socioeconomic scenario of the country, contributing to the generation of jobs, taxes, foreign exchange and income (Junior et al., 2012). In 2020, for example, companies in the segment were responsible for creating more than 536,000 direct jobs and 1.5 million indirect jobs, in addition to generating BRL 12.1 billion in federal taxes, which is equivalent to 0.9% of the country's total collection (IBÁ, 2021).

Due to its edaphoclimatic characteristics (soil and climate) and technological development achieved in the area of forestry, Brazil is highly competitive in the market for forest products (Juvenal; Mattos, 2002). However, increasingly globalized market pressures require companies to become increasingly competitive, providing high quality products at the lowest possible cost (Silva, 2001).

In the forestry sector, one of the most expensive steps is harvesting wood, which, together with transport, accounts for more than 40% of production costs (Machado; da Silva Lopes, 2000; Jaskiu, 2015; Chichorro et al., 2017). Due to the great impact of the harvest on the economic viability of forestry enterprises, it is essential to carry out efficient planning of this activity (Bredstrom et al., 2010; Fuentealba et al., 2019), since this allows reducing costs, increasing productivity, regulating the flow of wood and optimizing the use of machinery and equipment (Fidelis; Reis, 2019).

In general, harvest planning consists of determining the plots to be harvested in each period of time in order to meet the demand for wood, maximizing revenues (and/or minimizing costs) and satisfying any other restrictions then imposed (Werneburg, 2015; González-González et al., 2022). However, this can become a complex problem depending on the number of fields, due to the exponential increase in the number of variables (Gomide et al., 2009).

In solving this type of problem, traditional approaches, such as the use of linear programming (LP) and integer linear programming (ILP), have offered good results for the treatment of small dimensional problems (Nobre, 1999). However, for major

problems, the computational time spent often makes the use of these methods technically unfeasible (Gomide et al., 2009). Since obtaining information with quality and speed is considered essential for the success of decision-making in forest management, heuristics and metaheuristics become interesting alternatives in these situations, as they make it possible to obtain satisfactory answers with a reasonable processing time (Pereira, 2004; Gomide et al., 2009).

Several studies have been carried out using metaheuristics to solve forestry planning related problems. The Clonal Selection Algorithm method (CSA) was used by Júnior et al. (2018b) in this type of problem and presented good results when compared to ILP. However, the authors point out that the choice of parameters is essential for this, since invalid solutions may be obtained depending on the configuration used. The same could be observed in the works developed by Ferreira et al. (2018), Júnior et al. (2018a) and Matos et al. (2019) using Simulated Annealing (SA), Variable Neighborhood Search (VNS) and Genetic Algorithm (GA), respectively.

In addition to the choice of parameters, the strategy used to generate initial solutions is another factor that can impact the performance of such methods. In the mentioned works (Ferreira et al. (2018), Júnior et al. (2018a), and Matos et al. (2019)), the initial solutions, submitted to the optimization methods, were randomly generated. However, given the complexity of such issues, the large number of possible solutions, and any restrictions imposed, there is a high probability this strategy generates low quality solutions, which can impact the performance of optimization methods. For that matter, an alternative is using other optimization methods to generate initial solutions. In the work carried out by Júnior et al. (2021) the inclusion, in the initial population, of a solution obtained by rounding the optimal solution of the LP for the relaxed problem was tested as an alternative to increase the performance of GA, CSA, VNS and SA. In this work, authors observed a substantial improvement in the performance of GA, CSA and VNS algorithms; however, no positive effects were observed for the SA metaheuristic.

Considering the importance of obtaining good quality solutions quickly for successful forestry enterprises, this work proposes the development of a heuristic for generating valid solutions for a production planning-related issue, aiming at providing solutions that meet the restrictions imposed by the planner with less dependence on

parameters and low computational cost. Furthermore, considering the evidence that incorporating high-quality solutions into the initial population of metaheuristics can enhance their performance, we assess the impact of including varying quantities of solutions generated by the proposed heuristic in the initial solution sets of GA, CSA, VNS and SA. The results obtained showed that the proposed heuristic can be used as a viable alternative for quick decision making in the forestry planning process. In addition, tests showed that the inclusion of valid solutions in the initial population of GA and CSA can generate a significant improvement in the performance of these metaheuristics.

Materials and methods

In this section, the definition of the studied problem is carried out, as well as description of the heuristic developed and the metaheuristics Genetic Algorithm (GA), Clonal Selection Algorithm (CSA), Variable Neighborhood Search (VNS) and Simulated Annealing (SA) implemented and, finally, the presentation of those methods used for evaluations.

Definition of the problem

In this work, the database presented by Matos et al. (2019) referring to a eucalyptus forest (Eucalyptus) with ages varying between 1 and 6 years old and a total area equal to 4,210 hectares was used. According to the authors, this forest was divided into 120 stands (management units) and 81 different prescriptions (management alternatives) were generated for each stand, defining the sequence of harvesting and planting to be carried out during a planning horizon of 16 years. The definition of prescriptions took into account the cutting of management units aged between 5 and 7 years old and immediate planting after cutting.

Assigning a certain prescription to a stand defines the age of the forest stand in each period of the planning horizon, each period being equal to one year. The volu-

me of wood produced and related silvicultural costs vary according to the age of the stand. In this sense, the choice of prescription affects not only when the harvest will be carried out, but also the amount of wood harvested and the Net Present Value (NPV) obtained. The production and NPV values corresponding to the application of each management alternative in each stand are contained in the database, composed of 20 columns and 9720 rows. Columns 1, 2 and 3 contain the stand number, current age and applied prescription, respectively. Columns 4 to 19, the volumetric production of

	1	2	3	4	...	19	20
1	1	1	1	0	...	0	14703
2	1	1	2	0	...	0	14703
...	...	...	...	...	...	...	...
9719	120	6	80	8960	...	8960	442508
9720	120	6	81	8960	...	8960	442508

wood for each year of the planning horizon and column 20, the NPV (Figure 1).

Figure 1 - Database representation. The columns contain the number of the plot, age of the forest stand, applied prescription, production in each year of the planning horizon and NPV.

Font: Author (2022).

The forestry production planning problem studied consists of determining which of the 81 prescriptions should be assigned to each of the 120 stands to maximize the NPV and meet the minimum and maximum annual demand for wood. For this, each solution was represented as a vector with 120 positions, where each position can assume an integer value between 1 and 81. In this representation, the positions correspond to the stands and the assigned values, to prescriptions. The mathematical optimization model used considered the structure proposed by Johnson and Scheurman (1977), represented

$$MAX NPV = \sum_{i=1}^M \sum_{j=1}^P NPV_{ij} \cdot X_{ij} \quad (1)$$

Subject to:

$$\sum_{j=1}^P X_{ij} = 1 \quad \forall i \in \{1, \dots, M\} \quad (2)$$

$$\sum_{i=1}^M \sum_{j=1}^P V_{ijk} \cdot X_{ij} \geq Dmin_k \quad \forall k \in \{1, 2, \dots, 16\} \quad (3)$$

$$\sum_{i=1}^M \sum_{j=1}^P V_{ijk} \cdot X_{ij} \leq Dmax_k \quad \forall k \in \{1, 2, \dots, 16\} \quad (4)$$

$$X_{ij} \in \{0, 1\} \quad (5)$$

by Equation 1.

Where: $-NPV_{ij}$ is the NPV of the management unit i according to the management regime j ; X_{ij} is the binary decision variable, receiving value 1 when the management alternative j is assigned to the management unit i or 0; otherwise; M is the total number of management units; P is the total number of management alternatives for the management unit i ; V_{ijk} is the total volume of wood (m³) of the management unit i according to the management regime j in a period k of the planning horizon; $-Dmin_k$ and $-Dmax_k$ are, respectively, the minimum and maximum wood demand (m³) k in the planning horizon period.

The objective function (Eq. 1) seeks to maximize the financial return and is subject to binary (Eq. 5), integrity (Eq. 2) and production (Eq. 3 and Eq. 4) constraints. The integrity constraint guarantees the choice of only one prescription for each management unit, and the production constraints ensure the supply of the forest product at the desired levels in each period of the planning horizon (Pereira, 2004).

In this work, the objective function took into account the method of penalties,

reducing the calculation value of the function (Eq. 1) by R\$ 500,00 (five hundred Brazilian Reals) per volume of wood produced, exceeding or lacking in relation to D_{min} e D_{max} . The minimum and maximum wood demands stipulated by Matos et al. (2019)

$$Obj = \sum_{i=1}^n \sum_{j=1}^p NPV_{ij} X_{ij} - P \sum_{k=1}^T dk \quad (6)$$

are 140,000 m³ and 160,000 m³, respectively.

Where: P represents the penalty (R\$); dk is the absolute value of the volumetric deviation of wood in m³ in a period k of the planning horizon.

The dk value is calculated through the absolute difference between the total production of the solution and D_{min} e D_{max} . In order to obtain the total production, the amount of wood harvested in each year of the planning horizon in each stand is searched in the database and stored in the production history. Finally, the sum of the production history is carried out for determining total production.

Figure 2 illustrates the process used to calculate the total production, in which (a) represents a solution, (b) database, (c) production history and (d) total production. The solution has prescription 66 assigned to the first stand; therefore, the row containing number 1 in the first column, referring to the stand, and the number 66 (prescription) in the third column is spotted in the database. From the searched row, the values contained in columns 4 to 19 (production in each year of the planning horizon) are stored in the production history (c). This process is carried out for all stands and, subsequently, the sum of the production history is performed, so that total production (d) is obtained. From total production, it is possible to identify whether a given year of the planning horizon has production outside the established limits and calculate the



Figure 2 - Representation of the production calculation, where (a) represents a solution, (b) database, (c) production history, and (d) total production. In (d), the years in red (1, 2, 14, 15 and 16) had production outside the established limits, so for each cubic meter of wood in excess or lack of these, the defined penalty applies.
Font: Author (2022).

absolute value of the wood volumetric deviation.

Proposed approach

Description of the proposed heuristic

The proposed heuristic aims at finding valid initial solutions for the studied forestry planning problem, a solution considered valid when all years of the planning horizon have production within the minimum and maximum limits established. For this, the algorithm developed was based on the idea of defining, at each iteration, the years with the highest and lowest production, and seeking to reduce such differences by changing prescriptions in certain stands.

Algorithm 1 presents the proposed heuristic, in which the variable numSolutions is the number of valid solutions to be generated, DB is the database, and DBr is the re-organized database (detailed in section 2.2.1). The generated solutions are stored in the variable setSolutions. StopCritis the maximum limit of objective function calculations per run, PrunningCrit is the interval at which the pruning function will be performed.

Algorithm 1 [setSolutions] = Initial heuristic (numSolutions, DB, DBr)

```

1. Constants: StopCrit = 1000; PruningCrit = 16; Dmin = 140.000; Dmax =
   160.000; NumPrescriptions = 81; NumStands = 120;

2. for i = 1:numSolutions do
3.   StopCounter = 0;
4.   PruningCounter = 0;
5.   [solution] = Generate a random solution (NumStands, NumPrescriptions);
6.   [prodHist, totalProd] = Calculate production (solution, DB, Dmin, Dmax,
     NumStands);
7.   [numIrregulars] = Define the number of years with irregular production
     (totalProd, Dmin, Dmax);
8.   while numIrregulars > 0 do
9.     if StopCounter >= StopCrit then
10.      Break;
11.    else
12.      [lowYear, highYear] = Select the years (totalProd);
13.      [stand] = Choose a stand (prodHist, lowYear, highYear);
14.      [prescription] = Define the new prescription for the selected stand
        (stand, lowYear, highYear, DBr);
15.      [solutionaux] = Modify the stand prescription (solution, stand, pre-
        scription);
16.      [prodHistaux, totalProdaux] = Calculate the production of the new
        solution (solutionaux, DB, Dmin, Dmax, NumStands);
17.      [numIrregularsaux] = Define the number of years with irregular pro-
        duction (totalProdaux, Dmin, Dmax);
18.      if numIrregularsaux <= numIrregulars then
19.        solution = solutionaux;
20.        prodHist = prodHistaux;
21.        totalProd = totalProdaux;
22.        numIrregulars = numIrregularsaux;
23.      end if
24.      StopCounter = StopCounter + 1;
25.      PruningCounter = PruningCounter + 1;
26.      if PruningCounter >= PruningCrit and numIrregulars > 0 then
27.        [solution] = perform the pruning (solution, prodHist, totalProd,
          NumStands, NumPrescriptions, Dmin, Dmax);
28.        [prodHist, totalProd] = Calculate production (solution, DB, Dmin,
          Dmax, NumStands);
29.        [numIrregulars] = Define the number of years with irregular pro-
          duction (totalProd, Dmin, Dmax);
30.        PruningCounter = 0;
31.      end if
32.    end if
33.  end while
34.  setSolutions(i, :) = solution;
35. end for

```

D_{min} and D_{max} the defined minimum and maximum limits of annual production, the NumPrescriptions and NumStands represent the total number of prescriptions and plots, respectively.

To obtain a valid solution, the proposed heuristic generates a random solution and calculates the production, storing the production history and total production. With total production, it is possible to determine the number of years with irregular production (less than 140,000 m³ or greater than 160,000 m³). If all years have regular production, the solution is saved and the heuristic is closed. As long as this criterion is not met or until the maximum number of objective function calculations is exceeded, the algorithm sorts the production of all years and selects those with the highest and lowest production. The year with the highest production is stored in variable Y_{max} and the one with the lowest production, in Y_{min} .

Variables Y_{max} and production history are used to select a stand to have the prescription modified. For this, the criterion used is that, if possible, the plot selected under the current prescription has production greater than zero in Y_{max} and equal to zero in Y_{min} . These requirements were defined taking into account that changing the prescription can modify the years in which the plot will be harvested and, therefore, the total production of the solution. In this sense, the first requirement was adopted in order to increase the probability that the modification will reduce production in Y_{min} , and the second requirement is intended to prevent this modification from also reducing production in Y_{max} .

Figure 3 illustrates the selection of stands subject to modification in a solution that, as shown in the total production, has as Y_{max} (highlighted in red) and Y_{min} (highlighted in blue) the first and thirteenth years of the planning horizon, respectively. In the first column of the production history, which consists of the production of all stands in Y_{min} , stands 118, 119 and 120 are selected because their productions are greater than zero. Subsequently,

	1	2	3	4	5	...	12	13	14	15	16
1	0	0	0	0	0	...	0	0	5632	0	0
2	0	0	0	0	0	...	6468	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...
118	8190	0	0	0	0	...	8190	0	0	0	0
119	4140	0	0	0	0	...	0	0	0	0	0
120	8190	0	0	0	0	...	0	8960	0	0	0
Total Production	175542	131494	172975	129274	130838	...	153757	124962	165986	141769	130159

evaluating in column 13, stand 120 is excluded from the alternatives because it has a production different from zero in . Therefore, 118 and 119 would be alternative stands to be modified.

Figure 3 - Representation of production history and selection of plots subject to modification
Font: Author (2022).

The choice of the new prescription is carried out aiming at increasing the production of , maintaining the production of . For this, a reorganization of the database was proposed, evaluating, for each year of the planning horizon, which prescriptions could be assigned to each of the stands so that the harvest could occur in that year. For example, if stand 118 of the solution shown in Figure 3, were selected for alteration, the reorganized database would apply to define which prescriptions could be assigned to this plot so that there would be a harvest in the thirteenth year () and not in the first one ().

The method used to reorganize the database was to filter, for each stand and year of the planning horizon, prescriptions with production other than zero and copy them to a new file. Figure 4 shows the reorganized database, in which the first column refers to the stand and the second to the year of the planning horizon. The remaining columns

	Stand	Year	Prescriptions					
1	1	1	0	0	...	0	...	0
2	2	1	0	0	...	0	...	0
...	...	...	...	...	...	...	...	...
118	118	1	1	2	...	24	...	54
...	...	...	...	...	...	...	...	...
1558	118	13	7	8	...	66	...	0
...	...	...	...	...	...	...	...	...
1919	119	16	28	79	...	0	...	0
1920	120	16	28	79	...	0	...	0
	1	2	3	4		26		56

Figure 4 - Database reorganization. Prescriptions stored in columns 3 to 56, if assigned to column 1 plot, will generate harvest in column 2 year. In the example mentioned, prescriptions 7, 8 and 66 could be assigned to plot 118 (row 1558), so that this plot would have production in the thirteenth year. However, the application of prescriptions between 1 and 54 (row 118) generates a harvest in the first year. In this situation, an alternative would be prescription 66, as if applied it would generate a harvest in and not in.

Font: Author (2022).

contain the prescriptions that, if assigned to a stand, will generate production in the year in question. A zero value is assigned when no prescription will result in a harvest during the year under evaluation.

The prescription to be assigned to the defined stand is selected by searching the reorganized database for alternatives that result in harvest in the . Of these alternatives, those that generate harvest in the are excluded. Then, a prescription is randomly selected from the remaining options. If no prescriptions meet the criteria or all are equal to zero, the selection is made randomly among all prescriptions.

By changing the prescription in the chosen stand, the calculation of the number of years with irregular production is performed again. If the modified solution has an equal or lesser number of violations, it becomes the current solution, otherwise it is discarded. This process runs until the algorithm reaches the maximum number of objective function calculations or the solution has no years with uneven production.

In the proposed heuristic, the choice of stand and prescription were developed with the objective of performing a local search, in order to improve the current solution.

However, to prevent the algorithm from getting stuck in local optima, it was necessary to implement the pruning function in order to make changes in the neighborhood. The pruning function applies when the algorithm executes the described procedure PruningCrit 16 times and the solution remain invalid. This function resets the production and assigns random prescriptions to the plots that have harvested in years with excess production until total production of these years approaches . If there are no years with excess production, the function uses the year with the highest production.

Description of metaheuristics

The implemented metaheuristics Genetic Algorithm (GA), Clonal Selection Algorithm (CSA), Variable Neighborhood Search (VNS), and Simulated Annealing (SA) followed the definitions proposed by Gaspar-Cunha et al. (2012), Souza and Romero (2014), Júnior et al. (2018a), Rodrigues et al. (2004), respectively. On GA and CSA, all tests were performed using 50 initial solutions. These solutions were generated randomly or using the developed heuristic. The number of solutions generated by the heuristic is defined by the variable (Algorithm 1). The remaining amount, to reach 50 solutions, is generated randomly. The Variable Neighborhood Search (VNS) and Simulated Annealing (SA) metaheuristics use only a current solution that, in this work was randomly generated or through the proposed heuristic ($= 1$).

The GA uses the tournament selection method to select the parents for mating, and the mating of individuals was performed with a break point at a probability of 80%. The mutation carried out consisted of exchanging prescriptions of two randomly selected plots with a probability of 1%. This mutation was called swap and is shown in Figure 5. The tournament was also used as a method of replacing individuals in the

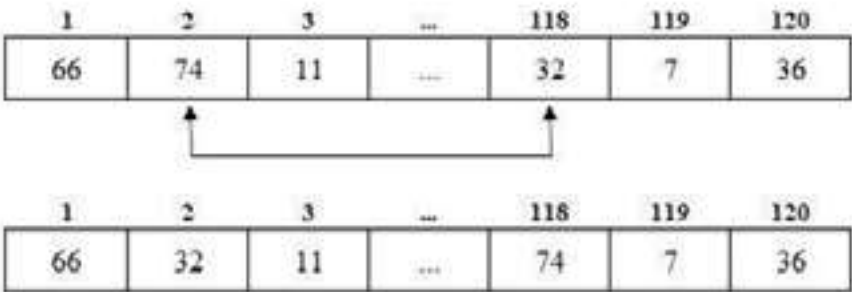


Figure 5 - Swap-type mutation in which prescriptions of two random plots are exchanged. The first and second images represent the solution before and after mutation, respectively.
Font: Author (2022).

population; however, elitism was adopted to avoid losing the best individuals. In the elitism used, the 10 best individuals among the parents and generated children are selected for the next generation.

In the Clonal Selection Algorithm (CSA), the selection rate for cloning adopted was 20% and the number of clones generated for each antibody was determined using Equation 7 (Batista, 2010). The generated clones were submitted to the hypermutation process, which consists of assigning random prescriptions to a number of plots inversely proportional to NPV (affinity). The mutation rate was obtained using Equations 8, 9 (Batista, 2010; Júnior et al., 2018b). The worst antibodies generated were replaced by the best clones in an elitist manner. Then, the antibody with the lowest affinity was replaced with a randomly generated solution in order to maintain diversity.

Where: where TC is the cloning rate (25%), N is the total amount of antibodies and round is the operator that rounds the value in parentheses to the nearest integer, TH is the hypermutation rate (10%), D* is the normalized NPV.

In VNS, the set of neighborhood structures implemented consisted of randomly modifying 1, 2, 3, and 4% of prescriptions (Júnior et al., 2018a). In SA, the initial temperature adopted was and the temperature reduction rate was 0.01 (Ferreira et al., 2018). In both cases, the Local Search procedure applied was swap, which involves exchanging the position of two random prescriptions. In VNS, 200 neighbors were used, while in SA, there were 30.

Assessment methods

To assess the performance of the proposed heuristic and the studied metaheuristics,

Scenario	Minimum demand (m ³)	Maximum demand (m ³)
1	140,000	160,000
2	140,000	150,000
3	150,000	160,000

Table 1 - Description of the evaluated scenarios. Scenario 1 contains the original constraints of the forestry planning problem and Scenarios 2 and 3 were generated by modifying the minimum and maximum production constraints, in order to evaluate the performance of the proposed heuristic

Font: Author (2022).

cs, two more restrictive scenarios were generated, in addition to the scenario described in section 2.1. The first scenario generated, called Scenario 2, has a minimum demand () equal to 140,000 m³ and a maximum demand () of 150,000 m³. The second scenario generated, Scenario 3, has same 150.000 m³ e 160.000 m³. As shown in Table 1.

The proposed heuristic was tested in each scenario, verifying the amount of objective function calculations necessary to obtain viable solutions. For this, 1,000 runs of the heuristic were performed, storing the number of calculations of the objective function, the solution obtained and the run time. Metaheuristics were implemented as presented adopting 5,000 objective function calculations as a stopping criterion, present in the works developed by Júnior et al. (2018a) and Matos et al. (2019). For each scenario, the proposed heuristic and metaheuristics were run 30 times, storing the maximum NPV and the number of valid solutions at the end of each run.

Subsequently, the performance of GA and CSA metaheuristics was evaluated using 1, 5, 10 and 15 individuals generated by the heuristic proposed in the set of initial solutions in the resolution of the first scenario studied (original). For VNS and SA, which uses only one current solution, the use of equal to 1 was evaluated. In these tests, the stopping criterion of 5,000 objective function calculations was adopted, considering the calculations used to generate the initial individuals, running 30 times for



each value of α and storing the maximum value of NPV obtained in each run. The results obtained by the metaheuristics using the randomly generated set of solutions and using the solutions generated by the proposed heuristics were submitted to statistical analysis at a significance level of 5%.

The implementation and experiments were performed using MATLAB R2020b, on a computer with an Intel® Core™ i5-7400 3.0 GHz processor with 4 GB RAM. For statistical analysis, Excel 2019 was used, initially verifying the assumptions for using the one-way analysis of variance (Anova one-way). The assumptions analyzed were normality, using the Shapiro-Wilk test, and homoscedasticity, applying the Levene test. In cases where these assumptions were met, Anova was performed, applying the Tukey mean comparison test when necessary. In cases where the assumptions were not met, the non-parametric Kruskal-Wallis test was used, followed by Dunn test.

Scenario	Average	Minimum	Maximum	Standard deviation	Valid solution
1	81.86	15	371	48.82	1000
2	1005.81	51	5000	761.66	999
3	5000	5000	5000	0	0

Table 2 - Evaluation of the proposed heuristic in the 3 scenarios in relation to the number of objective function calculations and the total number of valid solutions obtained
Font: Author (2022).

Result and discussion

Evaluation of the performance of heuristics and metaheuristics in the scenarios studied

Results of evaluation of the proposed heuristic in the 3 scenarios studied are shown in Table 2. For the first scenario (equal to 140.000 m– and 160.000 m–), an average of approximately 82 objective function calculations and 0.026 seconds were required to obtain a valid solution, finding solutions with production within the established limits in all runs performed. In the second scenario (equal to 140.000 m– and 150.000

m–), an average of 1,006 objective function calculations and 0.22 seconds were required, obtaining valid solutions in 99.9% of runs. While in scenario 3 (equal to 150.000 m– and 160.000 m–), no valid solutions were found in any of the 1,000 runs and the average time was 1.08 seconds.

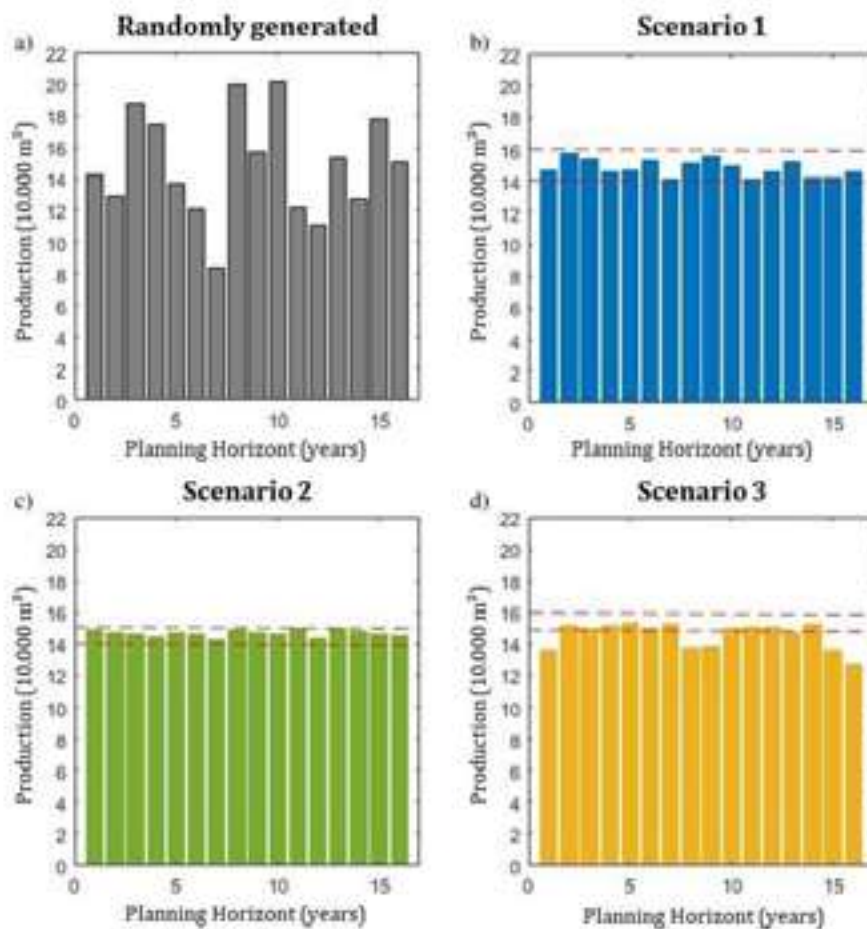


Figure 6 - Figure (a) illustrates the production over the 16-year planning horizon of a randomly generated solution. Figures (b), (c) and (d) were generated by the heuristic proposed in the first, second and third scenarios, respectively. Red lines represent the maximum and minimum demand for each scenario.

Font: Author (2022).

Figure 6 illustrates the production in each year of the planning horizon of 4 solutions, the first randomly generated and the others obtained by the proposed heuristic in each scenario. In the randomly generated solution, it is possible to observe a large variation in the volume of wood produced, with 83,682 m³ in the year with the lowest production, and 201,888 m³ in the year with the highest production. The solutions obtained by the heuristic proposed in the first and second scenarios present productions within the established limits. The solution produced by the heuristic in the third scenario is invalid, since the production of 6 out of 16 years of the planning horizon are below the stipulated minimum demand. However, this solution presents greater regularity when compared to the random one. The presented results indicate that the developed heuristic is able to generate valid solutions with a low computational cost in Scenarios 1 and 2. However, Scenario 3 proved to be more restrictive, with the possibility that there are no valid solutions for this problem under these conditions.

Then, the performance of the proposed heuristic and metaheuristics Genetic Algorithm (GA), Clonal Selection Algorithm (CSA), Variable Neighborhood Search (VNS) and Simulated Annealing (SA) in solving the 3 scenarios of the forestry planning problem were evaluated (Table 3). By checking the number of runs with valid solutions (at the end of the run), it is possible to observe that in the first scenario, contrary to the proposed heuristic, all the studied metaheuristics had at least one run with invalid solutions. These results demonstrate that obtaining solutions with production within the

Method	Average	Minimum	Maximum	Standard deviation	Execution with valid solutions
Scenario 1					
Heuristic	29990716	29593240	30392521	187132	30
GA	30431317	30153012	30681827	149990	29
CSA	30384715	30118383	30629176	129446	28
VNS	30560674	30357679	30819973	142232	29
SA	30755600	30497000	30970000	140324	28
Scenario 2					
Heuristic	30296563	30073647	30641133	131356	30
GA	28766777	24272351	30409291	1931754	6
CSA	28271868	24276469	30379693	1417596	2
VNS	28521090	21255287	30453983	1993373	3
SA	28632600	20030000	30737000	2520334	11
Scenario 3					
Heuristic	12558783	2173764	19499702	4539943	0
GA	15824019	8651377	21340680	3749860	0
CSA	17253809	11064146	22837152	2684917	0
VNS	15627888	6074459	21424824	3669773	0
SA	17046467	2288000	27674000	5152299	0

Table 3 - Results of the 3 metaheuristics in each scenario studied using equal to 0, containing the evaluated metaheuristic, the average, minimum and maximum Net Present Value, the standard deviation and verification of the existence of valid solutions at the end of the 30 runs.

Font: Author (2022).

established limits is not a trivial task in any of the scenarios, and indicate the potential of the heuristic for quick decision making in forest production planning. In the second scenario, the proposed heuristic, GA, CSA, VNS and SA obtained valid solutions in 100%, 20%, 6.66%, 10% and 36,7% of runs, respectively. In the third scenario, no valid solutions were found in any run.

Regarding the Net Present Value (NPV), it is important to highlight that the proposed heuristic does not aim to maximize. In this sense, the results presented in Table 3 do not refer to the best NPVs found by the heuristic, but to the NPVs of the solutions found at the end of the heuristic execution, which occurs when a valid solution is found or when the stopping criterion of 5,000 objective function calculations is reached. In the first scenario, the proposed heuristic required an average of 67 objective



function calculations to obtain valid solutions. These solutions had an average NPV of BRL 29,990,716.00, being the lowest among the methods evaluated. However, all metaheuristics required 5,000 function calculations.

Still in relation to Table 3, the proposed heuristic presented promising results for the second scenario. In addition to obtaining valid solutions in all runs, this method presented an average NPV greater than those of the evaluated metaheuristics, BRL 30,296,563.00, and a smaller standard deviation, BRL 131,356.00, using an average of 906 objective function calculations. In the third scenario, the proposed heuristic and the studied metaheuristics do not find valid solutions, demonstrating the complexity of the problem under these conditions.

Evaluation of the feasibility of applying the proposed heuristic for generating initial solutions

Average/ Standard deviation				
NumSolutions	GA	CSA	VNS	SA
0	30431317 (149990)	30384715 (129446)	30560674 (142232)	30755600 (140324)
1	30500953 (147202)	30427883 (157953)	30615555 (160255)	30760167 (128353)
5	30624375 (96735)	30575246 (112913)	.	.
10	30667067 (98946)	30582604 (96385)	.	.
15	30644007 (93582)	30585436 (103619)	.	.

Table 4 - Average Net Present Value (NPV) and standard deviation of the Genetic Algorithm (GA), Clonal Selection Algorithm (CSA), Variable Neighborhood Search (VNS) and Simulated Annealing (SA) in the first scenario studied using 0, 1, 5, 10 and 15 in the set of initial solutions.

Font: Author (2022).

Table 4 shows the comparison between results obtained by metaheuristics in solving the first (original) scenario of the forestry planning problem using different numbers of individuals generated by the proposed heuristic () in the set of initial solutions. Observing the results, it is possible to identify that the increase of valid solutions in the set of initial solutions generated an increase in the average NPV.

For better understanding the results, a statistical analysis of data presented in Table 4 was performed. For GA, considering the heterogeneity of variances verified by the Levene test (p – value = $2.02E-03$), the analysis was performed using the non-parametric Kruskal-Wallis test (p – value = $5.08E-10$), followed by the Dunn test. The Dunn test pointed out that there is no significant difference between the use of 0 and 1 initial individual. However, when increasing the amount of , it was observed that equal to 0 and 5, 10 and 15 have statistically significant differences, indicating that the use of the heuristic can impact the convergence of this metaheuristic when using larger amounts of initial individuals. The comparison between 1 initial individual and 5, 10 and 15 reinforces this result, showing that there is a significant difference between the use of 1 and a larger number of individuals. Finally, no significant differences were identified between the use of 5, 10 or 15 initial individuals.

In the evaluation of CSA, the one-way analysis of variance indicated the presence of at least one statistical difference between the groups (p – value = $1.47E-12$) and the Tukey test performed showed a result similar to that of the GA. In the VNS and SA, the

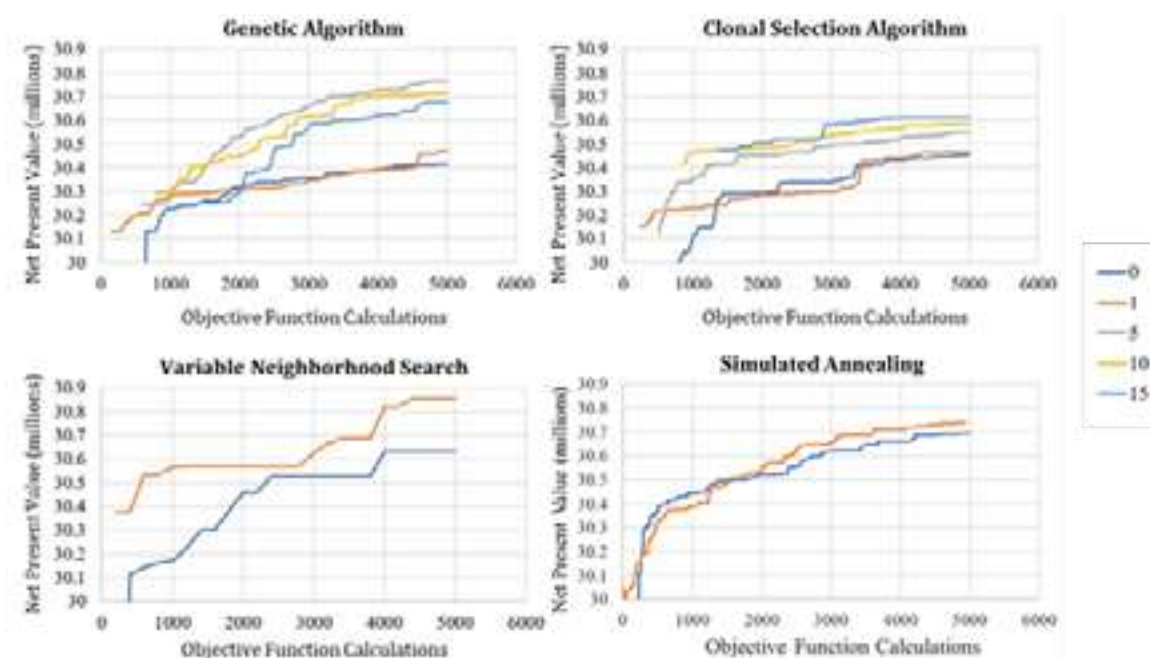


Figure 7 - Illustration of the convergence of the metaheuristics under study, using different amounts of . A run was performed for each amount of .

Font: Author (2022).

one-way analysis of variance revealed that there are no statistical differences between the groups (p – value = 0.1660, 0.8958). That is, the use of only one valid solution did not generate statistical difference in the result. These results can also be seen in Figure 7, which illustrates the NPV throughout the execution of the GA, CSA, VNS and SA metaheuristics with different amounts of .

In this figure it is possible to observe the consumption of the objective function calculations by the proposed heuristic, identified by the NPV starting point, and also the impact of the inclusion of valid solutions on the convergence of the metaheuristics. When using 0 or 1, it is possible to observe the approximation or overlap of the NPVs at some moments of the execution. With a greater number of valid solutions, a difference in convergence is visible, especially in GA.

Still in relation to Figure 7, the reduction in the NPV at the beginning of the SA execution, equal to 1, indicates that the solution generated by the heuristic was replaced by another of lower quality, accepted due to the high temperature at the beginning of the optimization process. The same behavior of the SA method was described by

Júnior et al. (2021), concluding that the use of good initial solutions did not affect the performance of this meta-heuristic. However, the inclusion of the solution generated by rounding the optimal linear programming solution in the initial CSA, GA, and VNS solutions significantly impacted the performance of these algorithms, generating a 3.93% increase in the average NPV in CSA (algorithm that had the greatest impact on average NPV). This result disregards, however, the computational cost necessary to obtain the solution generated by the PL.

Considerations about the proposed heuristic and literature

The proposed heuristic was developed to generate good initial solutions for other optimization methods. However, the results presented by the literature (JÚNIOR et al., 2018a, b; MATOS et al., 2019) and the metaheuristics evaluated indicate the possibility of applying this method for rapid decision-making in planning forestry production. In the study carried out by Matos et al. (2019), for example, the solution found by the entire programming (branch and bound), after 96 hours of processing, presented an NPV equal to BRL 31,964,100 and the GA with 2.74 minutes (using 50 individuals and 100 generations), obtaining an average NPV of BRL 30,336,526. If compared with these results, the NPV of the proposed heuristic (obtained with an average time of 0.026 seconds) corresponds to 93.83% and 98.86%. Compared to SA (the studied meta-heuristic that showed better performance), the proposed heuristic obtained an NPV corresponding to 97.51% using an average of 1.34% of the objective function calculations. To illustrate, the same SA executed with the stopping criterion of 67 objective function calculations presents an average NPV of BRL 5,003,598.00.

Considering the ability to provide good quality, valid solutions with a low computational cost, the proposed heuristic can be an interesting alternative for making quick decisions. However, as this method does not aim to maximize the NPV, a longer processing time or more objective function calculations would not generate an improve-

ment in performance. This is because the heuristic generates random valid solutions. In this sense, in situations where the decision maker has more time available, the use of heuristics alone may not be interesting.

To maximize the NPV, the proposed heuristic can be used in conjunction with other optimization methods, such as GA and CSA, to improve its performance. However, the use of valid solutions at the beginning of the optimization process does not guarantee that the final solutions will also be valid. This occurs because meta-heuristics prioritize maximizing the NPV, and invalid solutions may have a high NPV depending on the penalty applied. Several works (Rodrigues et al., 2004; Gomide et al., 2009; Nascimento et al., 2012; Matos et al., 2019) use the penalty method to implement these restrictions, however, the penalty values differ between the authors. For example, in the work of Júnior et al. (2021), Gomide et al. (2009), and Nascimento et al. (2012) the fine for each cubic meter of wood outside the production range is BRL 100.00, BRL 500.00 and BRL 100,000.00, respectively. Although maintaining a constant flow of production allows meeting market demands and also promotes benefits such as the regular employment of labor and the balance between annual income and expenses (Davis, 1996; Rodrigues et al., 1998), the tolerance for production variation depends on the analysis conducted by the decision maker.

Conclusion

In this work, it was proposed the development of a heuristic to obtain valid solutions for a forestry planning problem. The performance of the proposed heuristic was evaluated in 3 scenarios and compared with the GA, CSA, VNS and SA meta-heuristics. Then, different amounts of solutions generated by the proposed heuristic were inserted into the set of initial solutions of the studied metaheuristics and the impact on the performance of these methods was evaluated. Results indicate that the developed heuristic can be a viable alternative for quick decision making in forestry production planning, since it obtained valid solutions in 100% of the runs of the first scenario (original) and in 99.9% of the second scenario, demanding an average of 82 and 1006 ob-

jective function calculations, respectively. In addition, the results also indicate that the solutions generated by this heuristic can be used in the initial population of metaheuristics such as the Genetic Algorithm (GA) and Clonal Selection Algorithm (CSA), in order to improve their performance.

In future work, we intend to compare the impact of including solutions generated by the proposed heuristic and the relaxed PL in the set of initial solutions of the GA and CSA meta-heuristics.

Acknowledgment

The authors are grateful to Fundação de Amparo à Pesquisa do Estado de Minas Gerais - FAPEMIG for financial support.

References

- BATISTA, Lucas de Souza,. *Investigação de novas abordagens em sistemas imunes artificiais para otimização*. Dissertação (Mestrado em Engenharia Elétrica) —Universidade Federal de Minas Gerais, 2010..
- BREDSTROM, David; JÖNSSON, Petrus; RÖNNQVIST, Mikael. *Annual planning of harvesting resources in the forest industry*. *International transactions in operational research*, v. 17, n. 2, p. 155-177, 2010. DOI: <https://doi.org/10.1111/j.1475-3995.2009.00749.x>
- CHICHORRO, José Franklim; DA SILVA, André Luiz Pinheiro; ANDRADE, Wendel Sandro de Paula; HEGEDUS, Clovis Eduardo Nunes; KUBOYAMA, Filipe Quirino Aquira. *Custos e índices econômicos de povoamentos de eucalipto do Programa Produtor Florestal no Espírito Santo*. *Pesquisa Florestal Brasileira*, v. 37, n. 92, p. 447-456, 2017. doi: <https://doi.org/10.4336/2017.pfb.37.92.1232>
- DAVIS, K. P. *Forest management: regulation and valuation*. 2.ed. New York: McGraw-Hill, 1966. 519 p.
- GASPAR-CUNHA, António; TAKAHASHI, Ricardo; ANTUNES, Carlos Henggeler. *Manual de computação evolutiva e metaheurística*. Imprensa da Universidade de Coimbra, Coimbra, 2012. DOI: <http://dx.doi.org/10.14195/978-989-26-0583-8>.
- DE CASTRO, L. Nunes; VON ZUBEN, Fernando J. *The clonal selection algorithm with engineering applications*. *Proceedings of GECCO*. 2000. p. 36-39.

FERREIRA, Paulo Henrique Batista; MATOS, Lisandra Maria; DE ASSIS, Adriana Leandra; CABACINHA, Christian Dias; JÚNIOR, Carlos Alberto Araújo. *Influência dos parâmetros da metaheurística simulated annealing em um problema de planejamento da produção florestal*. Caderno de Ciências Agrárias-UFMG, 2018.

FIDELIS, Alessandro Cruvinel; REIS, Ricardo Pereira. *Decisão no planejamento bianual da colheita florestal: o uso da programação inteira mista*. Revista de Economia e Sociologia Rural, v. 39, n. 2, p. 9-26, 2019.

FUENTEALBA, Sebastián; PRADENAS, Lorena; LINFATI, Rodrigo; FERLAND, Jacques A.. *Forest harvest and sawmills: An integrated tactical planning model*. Computers and Electronics in Agriculture, v. 156, p. 275-281, 2019. DOI: <https://doi.org/10.1016/j.compag.2018.11.011>.

GOMIDE, Lucas Rezende; ARCE, Julio Eduardo; DA SILVA, Arinei Carlos Lindbeck. *Uso do algoritmo genético no planejamento florestal considerando seus operadores de seleção*. Cerne, v. 15, n. 4, p. 460-467, 2009.

GONZÁLEZ-GONZÁLEZ, José Mario; VÁZQUEZ-MÉNDEZ, Miguel Ernesto; DIÉGUEZ-ARANDA, Ulises. *Multi-objective models for the forest harvest scheduling problem in a continuous-time framework*. Forest Policy and Economics, v. 136, p. 102687, 2022. doi: <https://doi.org/10.1016/j.forpol.2021.102687>.

IBÁ. *Indústria Brasileira de Árvores. Relatório Anual 2020*. Disponível em: <https://iba.org/publicacoes/relatorios>; 2021. Acesso em: 31 dez. 2021.

JASKIU, Eldemar. *Planejamento florestal espacial atendendo exigências de áreas máximas e mínimas contínuas de colheita*. 63 p. Dissertação (Mestrado em Ciências Florestais) —Universidade Federal do Paraná, Curitiba, 2015.

JOHNSON, K. Norman; SCHEURMAN, H. Lynn. *Techniques for prescribing optimal timber harvest and investment under different objectives-discussion and synthesis*. Forest Science, v. 23, n. suppl_1, p. a0001-z0001, 1977. DOI: <https://doi.org/10.1093/forestscience/23.s1.a0001>.

JUNIOR, Antonio Almeida de Barros; DA SILVA, Gilson Fernandes; PEREIRA, Gustavo Willam; MELO, Robson. *Aplicações de heurísticas para a solução de um problema de planejamento florestal.: Simpósio Brasileiro de Pesquisa Operacional*. Rio de Janeiro, Brazil; 2012. p. 2430-241.

JÚNIOR, Carlos Alberto Araújo; CASTRO, Renato Vinícius Oliveira; MENDES, João Batista; LEITE, Helio Garcia. *Can linear programming assist metaheuristics in forest production planning problem?* FLORESTA, v. 51, n. 3, p. 751-759, 2021. DOI: <http://dx.doi.org/10.5380/rf.v51i3.72612>.

JÚNIOR, Carlos Alberto Araújo; MENDES, João Batista; CABACINHA, Christian Dias; DE ASSIS, Adriana Leandra; MATOS, Lisandra Maria Alves; LEITE, Helio Garcia. *Meta-heuristic clonal selection algorithm for optimization of forest planning*. Revista Árvore, v. 41, 2018b. DOI: <https://doi.org/10.1590/1806-90882017000600007>.

JÚNIOR, Carlos Alberto Araújo; MENDES, João Batista; DE ASSIS, Adriana Leandra; CABACINHA, Christian Dias; STOCKS, Jonathan James; DA SILVA, Liniker Fernandes; LEITE, Helio Garcia. *Tuning of the metaheuristic variable neighborhood search for a forest planning problem*. Cerne, v. 24, p. 259-268, 2018a. DOI: <https://doi.org/10.1590/01047760201824032538>.

JUVENAL, Thais Linhares; MATTOS, René Luiz Grion. O setor florestal no Brasil e a importância do reflorestamento. Banco Nacional de Desenvolvimento Econômico e Social. Acesso em: <http://web.bndes.gov.br/bib/jspui/handle/1408/3142>; 2002. Acesso em: 15 nov de 2021.

MACHADO, Carlos Cardoso; DA SILVA LOPES, Eduardo. Análise da influência do comprimento de toras de eucalipto na produtividade e custo da colheita e transporte florestal. *Cerne*, v. 6, n. 2, p. 124-129, 2000.

MATOS, Lisandra Maria Alves; JÚNIOR, Carlos Alberto Araújo; DE ASSIS, Adriana Leandra; CABACINHA, Christian Dias; FERREIRA, Paulo Henrique Batista; MAGALHÃES, Emanuelly Canabrava. Influência dos parâmetros da metaheurística algoritmo genético em um problema de planejamento florestal. *Advances in Forestry Science*, v. 6, n. 4, p. 767-774, 2019. DOI: 10.34062/afs.v6i4.6197.

NASCIMENTO, Flavio Augusto Ferreira; DIAS, Andrea Nogueira; FILHO, Afonso Figueiredo; ARCE, Julio Eduardo; MIRANDA, Gabriel de Magalhães. Use of Metaheuristics particle swarm optimization in Forest planning. *Scientia Forestalis*, Vol.40 No.96 pp.557-565, 2012.

NOBRE, Silvana Ribeiro. A heurística da Razão-R aplicada a problemas de gestão florestal. Dissertação (Mestrado em Administração da Empresa Rural) — Universidade Federal de Lavras, Lavras, 1999.

PEREIRA, Gustavo Willam. Aplicação da técnica de reconhecimento simulado em problemas de planejamento florestal multiobjetivo. Dissertação (Mestrado em Ciência da Computação) — Universidade Federal de Minas Gerais, Belo Horizonte, 2004.

RODRIGUES, Flávio Lopes, LEITE, Helio Garcia, SANTOS, Heleno do Nascimento, SOUZA, Agostinho Lopes, RIBEIRO, Carlos Antônio Álvares Soares. Metaheurística simulated annealing para solução de problemas de planejamento florestal com restrições de integridade. *Revista Árvore*, v. 28, p. 247-256, 2004. doi: <https://doi.org/10.1590/S0100-67622004000200011>.

RODRIGUES, Flávio Lopes, LEITE, Helio Garcia, SOUZA, Agostinho Lopes, RIBEIRO, Carlos Antônio Álvares Soares, SILVA, Márcio Lopes. Regulação de florestas equiâneas utilizando programação linear: uma aplicação da teoria do modelo II. *Revista Árvore*, v. 22, n. 2, p. 193-213, 1998.

SILVA, Gilson Fernandes da. Problemas no uso de programação matemática e simulação em regulação florestal. Tese (Mestrado em Engenharia) — Universidade de São Paulo, São Paulo, 2001.

SOUZA, Simone SF; ROMERO, Ruben. Algoritmo Imunológico Artificial CLONALG e Algoritmo Genético Aplicados ao Problema do Caixeiro Viajante. *Proceeding Series of the Brazilian Society of Computational and Applied Mathematics*, v. 2, n. 1, 2014. DOI: <https://doi.org/10.5540/03.2014.002.01.0106>.

WERNEBURG, Mauro Antônio Pereira. Planejamento em grandes empresas florestais no Brasil. Dissertação (Mestrado em Ciência Florestal) — Universidade Federal dos Vales do Jequitinhonha e Mucuri; Diamantina; 2015.



■ Giselle Beiguelman, frames do vídeo generativo da série Flora Rebellis, versão 2024.
Imagens geradas com Inteligência Artificial (Style Gans)

■ [Acesse o vídeo clicando na imagem](#)





MÁQUINAS COMPANHEIRAS:

Tecnologias para o cuidado, autonomia e interdependência

DOI: <https://doi.org/10.35699/2965-6931.2023.46904>

PAULA LEMOS VILAÇA FARIA*

resumo: Este ensaio pretende organizar e debater algumas ideias em torno do uso de tecnologias dentro do espaço doméstico como ferramentas no trabalho de cuidado. A partir da delimitação desse termo e suas latências a partir da pandemia de COVID-19, aqui se discute como essas “máquinas companheiras” têm sido assimiladas para trazer autonomia e acessibilidade para idosos e pessoas com deficiência, bem como o significado desses dois termos no contexto da sociedade capitalista em que essas tecnologias se inserem. Colocando em foco principalmente os recursos que utilizam inteligência artificial, é trazido para a discussão a forma como elas são concebidas e sua relação com o usuário/consumidor final, suspendendo a assimilação rápida desses recursos e propondo uma hesitação para se pensar a respeito de suas ambivalências.

PALAVRAS-CHAVE: domesticidade; autonomia; acessibilidade; tecnologia; economia do cuidado

Máquinas Compañeras: Tecnologías para el cuidado, la autonomía y la interdependencia

ABSTRACT: Este ensayo pretende organizar y discutir algunas ideas en torno al uso de las tecnologías dentro del espacio doméstico como herramientas en el trabajo de cuidado. A partir de la delimitación de este término y sus latencias a partir de la pandemia del COVID-19, discutimos aquí como estas “máquinas compañeras” han sido asimiladas para brindar autonomía y accesibilidad a personas mayores y personas con discapacidad, así como el significado de estos dos términos. en el contexto de la sociedad capitalista en la que se insertan estas tecnologías. Centrándonos principalmente en los recursos que utilizan inteligencia artificial, se discute la forma en que se conciben y su relación con el usuario/consumidor final, suspendiendo la rápida asimilación de estos recursos y proponiendo una vacilación para pensar en sus ambivalencias.

KEYWORDS domesticidad; autonomía; accesibilidad; tecnología; economía del cuidado.

* Universidade Federal de Minas Gerais

¹ Referência ao título do texto “Máquinas companheiras” de 2023, de Giselle Beiguelman, que por sua vez referencia o livro “O manifesto das espécies companheiras: Cachorros, pessoas e alteridade significativa”, de 2021, de Donna Haraway.

² BEIGUELMAN, Giselle. Máquinas companheiras. Revista Morel #7, 2023.

³ Disponível em: <https://www.uol/noticias/especiais/assistentes-de-voz-x-feminismo.htm?utm_source=subscribe&utm_medium=email#mulheres-que-orientam> Acesso em: <06/06/2023>

⁴ Ibidem

⁵ Disponível em: <<https://outraspalavras.net/feminismos/pa->

Em obras de ficção científica, principalmente as produzidas no ocidente por autores do chamado norte global, a figura do robô costuma ser representada de forma oscilante entre a figura companheira que vive conosco em harmonia, em uma amigável posição de servidão, e entre o que se rebela, inconformado com esse papel e outras incumbências dada por seus criadores, humanos feitos de matéria orgânica. Essas “máquinas companheiras”¹ são definidas muitas vezes pela palavra robô, já centenária, que veio do checo “robota” e apareceu pela primeira vez em uma peça do escritor Karel Čapek em 1921, possuindo em seu contexto o significado de “trabalho compulsório”, derivando do eslávico que têm a conotação de servidão². Hoje, essa figura meio similar ao Frankenstein de Mary Shelley é feita de engrenagens mais modernas, e pode ser até mesmo descorporificada, composta de algoritmos e uma antropomorfa “inteligência artificial”. Ainda que nesses últimos exemplos os robôs não se assemelhem fisicamente a seres humanos como ocorre em várias obras de ficção que vão desde “Máquinas como eu: E gente como vocês” (2019), do escritor britânico Ian McEwan, ou em “O homem bicentenário” (1976), do escritor russo-estadunidense Isaac Asimov, essas máquinas tomam para si as características de um. Aqui será convencionado que os termos máquinas e robôs poderão designar essas duas modalidades, ainda que seja mais complexo materializar nesse ideal de robô aqueles que existem sem um corpo, apenas processados como informação em mega computadores. Para além da fantasia e imaginários de futuros dominados por homens e mulheres de lata e pele sintética, essas personagens assimiladas e construídas no chamado “mundo real” incorporam questões similares às das narrativas ficcionais, principalmente ao espelhar vestígios sociais em suas concepções, demarcando os contextos em que foram concebidas e muito do pensamento de seus criadores.

Na famosa animação de 1962 de William Hanna e Joseph Barbera, “Os Jetsons”, a família do futuro não vai muito longe do formato de família nuclear branca e heterossexual da década de 1960 dos Estados Unidos, colocada como um símbolo patriótico a partir do espaço doméstico. A personagem da mãe, Jane Jetson, apesar de todo o aparato tecnológico disponível em sua mansão futurista, segue como uma dona de casa realizando as tarefas domésticas, ainda que sob o cargo de patroa. Coabitando o mesmo espaço, há também a personagem da empregada-robô Rosie que faz a maior parte desse trabalho, nos moldes coloniais tradicionais que essa figura costuma ser

representada: Rosie, uma robô mulher que trabalha sem parar para a família Jetson, confinada em uma mansão flutuante de Orbit City.

Apesar dos mais de 60 anos que separam essa animação do presente, permanece uma demarcação do gênero feminino na maioria de robôs voltados para o espaço da casa e para atividades de cuidado, principalmente as assistentes virtuais utilizadas no cotidiano das pessoas³. A fonoaudióloga de mídia do UOL Claudia Cotes, pontua:

Pesquisas dizem que mulheres modulam a voz melhor que os homens. As empresas visam o lucro, então vão escolher as vozes que são mais aceitas por todos, para vender mais. Que isso [assistentes de voz mulheres] reforça o fato das mulheres estarem sempre a postos para ajudar, isso é verdade. Por isso a importância de deixar mais opções como a voz masculina. Vivemos ainda em uma sociedade machista, mas se dermos essas opções, esses padrões serão quebrados. (COTES, 2023)⁴

Essa aplicação enviesada de gênero em objetos produzidos sem inerentemente possuírem um em ferramentas tecnológicas se relaciona com o machismo e o racismo estrutural, envolvendo principalmente a chamada economia do cuidado. Ganhando evidência durante a pandemia de COVID-19 pelo aumento da carga de trabalho das mulheres confinadas em casa e tendo que majoritariamente realizar a manutenção da casa e cuidar de crianças, idosos e doentes, as separações entre trabalho produtivo e trabalho reprodutivo se dissiparam, tornando-o de mais difícil delimitação (Mezzadri, 2020). A economista Isabella Callegari define:

De forma resumida, a Economia do Cuidado pode ser dividida em (i) trabalhos diretos, que são aqueles diretamente relacionados à outra pessoa, como, por exemplo, dar comida, dar banho, ajudar no dever de casa, levar alguém ao médico; e (ii) trabalhos indiretos, aqueles necessários à manutenção de si e dos demais, mas que não constituem uma atividade direta com o outro, como cozinhar, lavar, limpar, ir ao mercado, e, no contexto rural, adicionam-se tarefas como abastecimento de água, fazer fogo, plantar e colher. É também estratificada em (i) trabalho não pago, que pode ocorrer em contexto doméstico ou no voluntariado; e (ii) trabalho pago, que pode ser formal ou informal, englobando os setores de saúde, serviço social e educação, bem como o de trabalhadoras/es domésticos e o de cuidados pessoais. Os cuidados podem ser direcionados a crianças, idosos, pessoas com deficiência, doentes, temporariamente desabilitados e ao autocuidado. (Callegari, 2021)⁵

A filósofa italiana Silvia Federici possui inúmeros estudos relacionados ao trabalho reprodutivo e de cuidado e sua relação com o capital a partir de uma leitura feminista de Marx. Em seu texto de 2015, “Reencantando o mundo: tecnologia, corpo e construção dos comuns”, ela traz o uso de tecnologia para a discussão, destacando sua ausência de neutralidade e como, mesmo com as tecnoutopias da década de 1950 e 1960, sendo “Os Jetsons” um exemplo, sua implementação não reduziu o trabalho realizado. A autora diz:

No entanto, é difícil deixar de acreditar que a introdução do computador tenha sido um benefício para a humanidade, que reduziu a quantidade de trabalho socialmente necessário e aumentou os bens sociais e a capacidade de cooperação. Entretanto, quando calculamos as consequências da informatização, fica difícil ter uma visão otimista da revolução da informação e da sociedade baseada no conhecimento. A informatização também aumentou a capacidade militar da classe capitalista e sua vigilância sobre nosso trabalho e nossa vida - consequências que fazem os benefícios do uso de computadores pessoais perderem seu brilho (Mander, 1991). É importante notar que a informatização não reduziu nem mesmo a jornada semanal de trabalho, uma promessa de todas as tecnoutopias desde a década de 1950, nem o fardo do trabalho físico. Estamos trabalhando mais do que nunca.” (Federici, 2015, p.)

Porém, abre-se espaço para uma outra discussão, que torna essa afirmação um tanto problemática, principalmente quando a tecnologia se tornou parte essencial do cotidiano de muitas pessoas, ainda que tenha outras implicações, como a não redução da jornada de trabalho mencionada acima e indagações sobre quem de fato produz e quem consome essas ferramentas. Essa afirmação parece assumir uma posição um tanto generalista ao colocar a tecnologia na dicotomia de ser boa ou ruim, principalmente ao pontuar suas consequências. Mas Federici, em outro texto, também de 2015, “Da crise aos comuns: trabalho reprodutivo, trabalho afetivo, tecnologia e a transformação da vida cotidiana”, escreve de forma bem clara a forma como há um caráter especial no trabalho de cuidado no sentido de ser extremamente complexo e árduo, e como a emulação dele em tecnologias é bastante problemática, partindo de lugares demarcados e gerando consequências bastante complicadas:

O trabalho doméstico, incluindo o trabalho de cuidado e o trabalho afetivo, é extremamente segregado, realizado de uma maneira que nos separa, individualiza nos-

sos problemas e oculta nossas necessidades e sofrimentos. Também é extremamente árduo: exige a execução de muitas atividades, em geral simultâneas, que não podem ser mecanizadas e que são realizadas principalmente por mulheres, na forma de trabalho não remunerado, muitas vezes somado a uma atividade remunerada em período integral. Sem dúvida, a tecnologia – da comunicação, em particular – desempenha um papel na organização do trabalho doméstico e agora é parte essencial de nossa vida cotidiana. Mas, como argumenta Fortunati (1998), ela serviu principalmente para substituir a comunicação interpessoal, e não para torná-la melhor, permitindo que cada membro da família escape da crise comunicativa refugiando-se na máquina. Da mesma forma, os projetos de empresas japonesas e estadunidenses de robotizar nossa reprodução – por meio de robôs-enfermeiros (*nursebots*) e de robôs-companhia (*lovebots*) personalizados para satisfazer nossas necessidades (Folbre, 2006) – são mais um sinal do crescimento da solidão, da perda de relações de apoio, e menos uma alternativa propriamente dita a esses fenômenos; além disso, é difícil que esses robôs de fato estejam presentes em muitos lares no futuro. Por isso, é de suma importância o esforço das mulheres para, acima de tudo, desprivatizar nossa vida cotidiana e criar formas cooperativas de reprodução. Isso abre caminho para um mundo em que cuidar dos outros pode se tornar uma tarefa criativa, e não um fardo, e para quebrar o isolamento característico do processo de nossa reprodução, criando laços de solidariedade sem os quais a vida é um deserto afetivo no qual não temos poder social. (Federici, 2015, p. 270-271)

Além dos lovebots e nursebots mencionados por Federici, atualmente têm sido lançados outros dispositivos similares, com funções específicas para reproduzir esse trabalho de cuidar dos outros. O jornalista James Vincent, em um texto para a The Verge, descreve os robôs de companhia distribuídos para mais de 800 idosos pelo estado de Nova Iorque como uma “nova forma de lidar com a epidemia de solidão no ocidente”⁶. ElliQ, mais uma ferramenta com gênero delicadamente demarcado, “pode projetar empatia e formar laços com os usuários. O robô deve se lembrar de detalhes importantes sobre a vida de um usuário e moldar seu personagem de acordo com o dele. Ele/a fará mais piadas se o usuário tende a rir muito, por exemplo.” Mais adiante no texto, Vincent fala que o New York State Office for the Aging (Nysofa), responsável pela iniciativa, tem a intenção de identificar os idosos que se classificam como aptos a

6 Disponível em: <https://www.theverge.com/2022/5/25/23140936/ny-state-distribute-home-robot-companions-ny-sofa-elliq> Acesso em: <31/05/2023>

⁷ Cotação do dólar feita no dia 01/06/2023.

⁸ Site de apresentação do robô Loona e seus recursos: <https://keyirobot.com/pages/loonade-tail?utm_source=substack&utm_medium=email> Acesso em: <31/05/2023>

⁹ DIBENEDETTO, Chase. Be My Eyes meets GPT-4: Is AI the next frontier in accessibility? Mashable. Disponível em: <<https://mashable.com/article/chatgpt-openai-be-my-eyes-accessibility>> Acesso em <17/05/2023>

¹⁰ Ibidem

¹¹ Ibidem

¹² Disponível em: <<https://www.linkedin.com/feed/update/urn:li:activi>>

se beneficiarem do robô, sendo eles os que seguem o critério da idade de 75 anos ou mais, que tenham acesso a Wi-Fi e familiaridade com tecnologia e estejam isolados ou sozinhos. Os custos por fora da iniciativa de Nova Iorque, a partir da empresa desenvolvedora da ferramenta para se ter essa companhia e seu serviço possuem uma taxa inicial de 250 dólares (R\$1250), com uma taxa mensal de 30 dólares (R\$150)⁷.

Recentemente, uma parente não muito distante de ElliQ foi anunciada na edição de 2023 do SXSW. Loona⁸, “o robô companheiro mais inteligente” segundo seu slogan, já é destinado para toda a família. Com “corpo pequeno, personalidade grande”, “Loona pode entender milhares de instruções e fornecer vários comentários interessantes. Você pode jogar os jogos ChatGPT do Loona para várias interações, como jogos de RPG, dramatizações, aprendizado de idiomas e etc.” No vídeo de apresentação, como um cachorrinho bem treinado, “um pet incrível que por acaso é um robô”, Loona (apresentada pelo pronome “ela”) recebe seu dono ao chegar em casa, brinca com as crianças e com o cachorro de carne e osso, se alonga, late e recebe carinho na cabeça. Mas também faz o escaneamento de segurança da sua casa, que pode ser acessado remotamente “caso você esteja preocupado enquanto estiver fora”. Se observarmos atentamente, o vídeo poderia ser uma versão atualizada dos Jetsons.

Mas recentemente, o aplicativo Be My Eyes tem testado voluntários virtuais⁹, ferramentas dirigidas por inteligência artificial para auxiliar seus usuários. Inicialmente criado para pessoas se cadastrarem como “olhos” para outras pessoas com deficiência visual, é inegável a autonomia que a inteligência artificial dá a quem demanda esses mesmos olhos no dia a dia. Uma das voluntárias para os testes desse recurso no Be My Eyes, Lucy Edwards, diz que “estava pronta para a IA antes mesmo de ela existir, porque sabia o que estava perdendo. Toda a internet poderia mudar completamente para mim.”¹⁰ No geral, ela encara como bastante positiva a implementação dessa ferramenta no aplicativo, que possibilita uma maior autonomia, por ser “apenas ela e seu smartphone em harmonia e sem outra assistência”.¹¹ Em uma linha parecida de ampliar a autonomia, a estudante de Ciências da Computação na Vellore Institute of Technology Priyanjali Gupta desenvolveu um sistema de Inteligência Artificial que traduz a linguagem de sinais para o inglês. A ferramenta “aproveita uma interface aberta Tensorflow de detecção de movimentos para traduzir os gestos realizados pelas mãos e transferir a um software de IA pré-treinado intitulado *ssd_mobilenet*, que realiza uma leitura da linguagem de sinais para o inglês.”¹²

¹³ HAN, Byung-Chul, 2022, p. 47.

É inegável que esse tipo de tecnologia possui uma aplicabilidade e importância enormes, assim como outras propostas similares de uso de IA para ampliar autonomia e acessibilidade em algum nível. Mas cabe refletir que, essa captura de movimentos, frames e sensores não é exatamente o que se pode chamar de falar a língua do outro e compreendê-lo. Enquadra-se a linguagem, aqui trazendo a noção de Butler (2016), deixando a vida que vaza e excede às condições normativas do espaço da tela do lado de fora, emoldurando traduções mediadas pela inteligência artificial. Cria-se uma ferramenta para mediar a linguagem com o pressuposto da autonomia de forma que não seja necessário sair de seu lugar e sua linguagem própria para entender a do outro, porque esse outro perde de certa forma sua corporeidade ao ser transmitido por meio dos dispositivos utilizados. Esse tipo de interação a partir da tecnologia parece intencionalmente obliterar a si mesma por parecer uma mediadora descorporificada, mas é determinante de como as relações se desencadearão de acordo com as possibilidades que ela mesma oferece. Essas superfícies lisas através das quais se acessa esses recursos traz o que o filósofo coreano Byung Chul-Han em seu livro de 2022 “Não-coisas: reviravoltas do mundo e da vida” chama de ausência de negatividade ou de atrito, removendo algumas resistências da realidade material. A ausência desses obstáculos, desses atritos não-dóceis e até mesmo desagradáveis segundo Chul-Han, trazem a sensação de que “tudo está acessível e disponível”¹³. Colocar esse tipo de recurso como alternativa para a autonomia de determinados grupos, facilitando o diálogo com outros corpos considerados padrão na sociedade contemporânea, sem deficiência, parece criar uma ambivalência. O isolamento desses grupos com o recurso tecnológico parece ser superficialmente remediado, ainda que de um modo que de fato possa ser útil a eles, mas também aumentado, uma vez que essa ferramenta é uma camada de afastamento de um contato direto entre as pessoas.

Cabe o questionamento, retomando as reflexões de Silvia Federici: estamos delegando para a inteligência artificial implementada em dispositivos específicos ou em recursos já utilizados, uma tarefa coletiva de cuidar das outras pessoas? A tecnologia está realmente se propondo a de fato dar autonomia a grupos como pessoas com deficiência ou idosos, por exemplo, ou autonomia para que o trabalho de cuidado não tenha que ser realizado por outras pessoas? Ao colocar essas tecnologias nesse lugar de cuidado, ao mesmo tempo que parece uma espécie de solução para essas tarefas,

esse mesmo trabalho realizado principalmente por mulheres é obliterado como uma questão social a ser discutida em larga escala sob o viés de gênero e do capital, ainda que vestígios dessas demarcações permaneçam nessas ferramentas.

Além disso, a ideia de independência numa sociedade consumida por um mito neoliberal de realização pessoal pela autonomia é extremamente cruel quando se necessita de outras pessoas em maior ou menor grau. No texto da jornalista Vivian Whiteman, ao falar dessa valorização desse eu capaz de “se virar sozinho” pontua, de forma muito pertinente:

Se depender é tudo de ruim, então reina a tendência do “auto”. A autoestima, o autoamor ou amor-próprio, o autocuidado, uma certa ideia de unidade completa que não precisa absolutamente do outro, que chega para trocar ou para compartilhar estando, como se diz, ‘plena’. Não é esquisito isso? Não parece algo que se confunde com consumo, sabe, como se as pessoas não pudessem compartilhar indefinições e vulnerabilidades? Será que todo “auto” não passa pelo alter, pelo outro? Alguém aprende a amar sozinho? O próprio eu não é em boa parte outro? (Whiteman, 2022)

Quando a inteligência artificial é colocada nesse lugar de ser a ferramenta incumbida de dar autonomia a pessoas com deficiência e fazer companhia para idosos, é de extrema importância analisar mais a fundo para quem exatamente ela está concedendo essa suposta dádiva de não precisar dos outros. A hipótese discutida aqui é que se trata menos de um altruísmo e preocupação com o outro e cada vez mais de uma vontade de delegar o cuidado, a possibilidade de retirar da esfera social e coletiva o contato com pessoas que o demandam, dando autonomia principalmente para os criadores dessas tecnologias que no final das contas não precisarão, se já não precisam, lidar com o trabalho que é cuidar de outra pessoa. Além de partir para uma automação de um trabalho majoritariamente realizado por mulheres e invisibilizado principalmente por não ser remunerado, as discussões se atropelam na mesma medida em que se tornam um emaranhado cada vez mais denso. Parece que a criação de recursos tecnológicos para supostamente reduzir ou substituir o trabalho humano de cuidado está intencionalmente ligado à sua não discussão. Debates que poderiam estar inseridos em discussões de políticas públicas, como a solidão dos idosos em Nova York, recebem alternativas individualizadas e privadas, oferecidas principalmente a quem pode pagar por elas.

¹⁴ UNESCO. Recomendação sobre a Ética da Inteligência Artificial. UnesDoc Digital Library, 2021. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000381137_por> Acesso em: <13/05/2023>

Essa suposta autonomia está cada vez mais distante do que se entende por acessibilidade e da compreensão da economia do cuidado como uma peça essencial no funcionamento do capital. Apesar de supostamente concedê-la para os usuários, essa autonomia sem a participação de outras pessoas vai contra a interdependência de se viver coletivamente. Porque a acessibilidade implica que uma espécie de “mundo lá fora” seja construído de forma que acolha a todos, ou seja, um espaço de coletividade. E quanto mais a ênfase é colocada em uma independência dentro de uma sociedade capitalista, mas soa como se fosse uma necessidade de ser legitimado dentro dela, assim como Whiteman menciona, como um consumidor.

Aprovada em novembro de 2021, a Unesco produziu uma cartilha chamada “Recomendação sobre a Ética da Inteligência Artificial”¹⁴, que entre suas diretrizes, possui uma série de tópicos separados por áreas de ações políticas, sendo “gênero” e “saúde e bem-estar social” dois exemplos. No tópico 90, dentro da área política sobre gênero é recomendado que:

Os Estados-membros devem assegurar que os estereótipos de gênero e os vieses discriminatórios não sejam traduzidos em sistemas de IA; em vez disso, tais elementos devem ser identificados e reparados de forma proativa. São necessários esforços para evitar o efeito negativo crescente das divisões tecnológicas para alcançar a igualdade de gênero e evitar diversas formas de violência, como assédio, intimidação ou tráfico de meninas, mulheres e grupos sub-representados, inclusive no mundo online. (UNESCO, 2021, p. 33)

Posteriormente, no tópico 125 na área política de saúde e bem estar social, há a recomendação:

Os Estados-membros devem desenvolver diretrizes para as interações humanos-robôs e seu impacto sobre as relações humanos-humanos, com base em pesquisas e dirigidas ao futuro desenvolvimento de robôs, e com especial atenção à saúde mental e física dos seres humanos. Deve ser dada particular atenção ao uso de robôs na assistência médica e no cuidado de idosos e pessoas com deficiência, na educação, e robôs para uso de crianças, robôs de brinquedo, chatbots e robôs de companhia para crianças e adultos. Além disso, a assistência de tecnologias de IA deve ser aplicada para aumentar a segurança e o uso ergonômico de robôs, inclusive em um ambiente de trabalho que integre humanos e robôs. Também deve ser dada especial atenção para a possibilidade de se usar a IA para manipular e abusar de vieses cognitivos humanos. (UNESCO, 2021, p. 38)

¹⁵ Disponível em: <https://news.artnet.com/opinion/quantitative-aesthetics-2276351>> Acesso em: <31/05/2023>

¹⁶ Disponível em: https://www.linkedin.com/posts/think-olga_esgotadas-labthinkolga-economiadocuidado-activity-

Tendo como ponto de partida quem cria esse tipo de tecnologia, seu contexto e quais seus objetivos elas se propõem, já se sabe que essas recomendações não estão sendo cumpridas. Por já nascerem tomando o partido dos vieses de seus criadores e visando o lucro, assim como Giselle Beiguelman diz que o Chat GPT “já nasce embalado por uma série de problemas, como a pressão de mercado que forçou seu lançamento antes de estar pronto” (Beiguelman, 2023), é praticamente impossível dizer que essas ferramentas serão o futuro do cuidado e da autonomia. A chamada “falácia de McNamara” ou “falácia quantitativa”, segundo a qual o que não pode ser mensurado ou medido não é considerado importante, ocupa um papel muito relevante atualmente na logística cotidiana que lida com o digital.¹⁵ Apesar de existirem pesquisas estatísticas de horas dedicadas ao trabalho reprodutivo do cuidado ou ao trabalho doméstico que sobrecarrega e afeta quase metade das mulheres brasileiras¹⁶, por exemplo, o cuidado como trabalho em si não é algo exatamente mensurável por números e passível de ser traduzido para a linguagem atualmente utilizada pelas ferramentas guiadas por inteligência artificial. A acessibilidade também não se dá visando apenas a autonomia de PCDs como operadoras dessas tecnologias, mantidas ainda isoladas com seus dispositivos e restrita à quem pode tê-los.

A escritora de ficção científica Ursula K. Le Guin possui um ensaio chamado “A teoria da bolsa da ficção”, no qual discorre sobre como as narrativas, não só as ficcionais, mas também sociais, privilegiam imaginários que possuem uma lógica similar à de uma flecha ou de uma arma, possuindo a natureza de uma ferramenta que é objetiva e linear que tem um alvo determinado e um herói que a opera. Le Guin propõe então uma alternativa pela substituição da lógica da flecha por uma teoria determinada pela bolsa, dentro do qual “ainda há sementes para serem coletadas, e espaço na bolsa das estrelas” (LE GUIN, 2021. P. 24). Essa associação tanto à arma quanto à bolsa de Le Guin remete à atividade de caçadores e coletores na pré-história, que dividiam suas tarefas de modo que as mulheres eram responsáveis pela coleta, enquanto os homens pela caça. Ao voltar da caça com alguma criatura abatida, as histórias das desventuras para consegui-la eram de cunho heroico, na qual o caçador “vencia” ao final. Enquanto isso, as mulheres e suas bolsas carregavam mantimentos e objetos para inúmeros fins, um universo inteiro e não linear de recursos a partir de suas atividades. Essa associação da autora pode ser feita ao modo como estão sendo pensadas essas tecnologias aqui

17 Disponível em: <https://thaisnunes.substack.com/p/por-que-o-vale-do-silicio-deveria?utm_source=substack&utm_medium=email> Acesso em: <31/05/2023>

discutidas voltadas para o trabalho do cuidado. Sendo uma atividade reprodutiva e não linear, aplicar a lógica da flecha, de ferramentas com objetivos claros, parece incoerente. O cuidado e a interdependência mais se associam a ideia de uma bolsa dentro da qual “se cabe” de outras maneiras, fugindo dessa lógica de remediar questões consideradas inconvenientes ligadas a corpos dissidentes.

Máquinas podem sim até certo ponto serem companheiras na interdependência de humanos entre si, assim como serem auxiliares em várias tarefas de cuidado, mas desde que não sejam o argumento final como uma proposta de solução de problemas que não passam pela base das relações sociais e da estrutura capitalista do mundo contemporâneo. Com bases da filósofa Donna Haraway e seu “Manifesto Ciborgue”, a reflexão da escritora Thaís Nunes em seu texto “Por que o Vale do Silício deveria reler Mary Shelley?” vai exatamente ao ponto sobre essas ferramentas: “A responsabilidade não está apenas em criar uma IA mas também em cuidar, a intensa relação com o Outro, e esse ponto é essencial para repensarmos a tecnologia não apenas como sinônimo de capitalismo e progresso.”¹⁷

Referências

BEIGUELMAN, Giselle. *Máquinas companheiras*. Revista Morel #7, 2023.

BUTLER, Judith. *Quadros de guerra: quando a vida é passível de luto?* Rio de Janeiro: Civilização Brasileira, 2016.

CALLEGARI, Isabela. *Para entender a urgência da Economia do Cuidado*. Outras Palavras, Disponível em: https://outraspalavras.net/feminismos/para=-entender-a-urgencia-da-economia-do-cuidado/?fbclid=IwAR2SwujYUHcGSL_9TàK5GJoetSdWgriNRqckTàX-wlhJ7nu3BUUol5sE3t8. Acesso em: 14 dez. de 2021.

CHIANG, Ted. *Will AI become the new McKinsey?* The New Yorker. Disponível em: <https://www>.

newyorker.com/science/annals-of-artificial-intelligence/will-ai-become-the-new-mckinsey?utm_source=substack&utm_medium=email. Acesso em: 18 mai. de 2023.

DI BENEDETTO, Chase. *Be My Eyes meets GPT-4: Is AI the next frontier in accessibility?* Mashable. Disponível em: <https://mashable.com/article/chatgpt-openai-be-my-eyes-accessibility>. Acesso em: 16 mai. de 2023.

FEDERICI, Silvia. *Da crise aos comuns: trabalho reprodutivo, trabalho afetivo, tecnologia e a transformação da vida cotidiana* [2015]. In: Silvia Federici. *Reencantando o mundo: feminismo e a política dos comuns*. Tradução: Coletivo Sycorax – São Paulo: Elefante, 2022. pp. 254-271

FEDERICI, Silvia. *Reencantando o mundo: tecnologia, corpo e construção dos comuns* [2015]. In: Silvia Federici. *Reencantando o mundo: feminismo e a política dos comuns*. Tradução: Coletivo Sycorax – São Paulo: Elefante, 2022. pp. 272-285

HAN, Byung-Chul. *Não-coisas: reviravoltas do mundo da vida* / Byung Chul-Han. Tradução: Rafael Rodrigues Garcia. Petrópolis, RJ: Vozes, 2022.

HARAWAY, Donna. *Manifesto ciborgue: ciência, tecnologia e feminismo-socialista no final do século XX* / Donna Haraway. Tradução: Tomaz Tadeu. In: Tomaz Tadeu (org.) *Antropologia do ciborgue: as vertigens do pós-humano*. Belo Horizonte: Autêntica, 2000, pp. 37-129

HARAWAY, Donna. *O manifesto das espécies companheiras: cachorros, pessoas e alteridade significativa* / Donna Haraway. Tradução: Pê Moreira; revisão técnica e posfácio Fernando Silva e Silva. – 1. ed. – Rio de Janeiro: Bazar do Tempo, 2021.

LE GUIN, Ursula K. *A teoria da bolsa da ficção* / Ursula K. Le Guin. Tradução: Luciana Chieregati, Vivian Chieregati Costa; introdução de Juliana Fausto; posfácio de Luciana Chieregati. - São Paulo: n-1 edições, 2021.

LEVY, Nathália. #12/24: *Interdependência. Minha existência na internet*. Disponível em: <https://nathalialevy.substack.com/p/1224-interdependencia>. Acesso em: 16 mai. de 2023

MEZZADRI, Alessandra. *The informal labours of social reproduction*. *Global Labour Journal*, 2020, 11 (1).

NUNES, Thaís. Por que o Vale do Silício deveria reler Mary Shelley? *Thaís Literatura*. https://thaisnunes.substack.com/p/por-que-o-vale-do-silicio-deveria?utm_source=substack&utm_medium=email. Acesso em: 18 mai. de 2023.

TAYLOR, Sunny. *The Right Not to Work: Power and Disability*. *Monthly Review*. Disponível em: <https://monthlyreview.org/2004/03/01/the-right-not-to-work-power-and-disability/>. Acesso em: 16 mai. de 2023.

The Jetsons (Os Jetsons). Criação: William Hanna e Joseph Barbera. Estados Unidos: ABC Network. Distribuição no Brasil: SBT. Série original: 1962-1963 / Nova série: 1985-1987. Nfl de temporadas: 3. Nfl de episódios: 75.

VINCENT, James. *NY State is giving out hundreds of robots as companions for the elderly*. *The Verge*. Disponível em: <https://www.theverge.com/2022/5/25/23140936/ny-state-distribute-home-robot-companions-nysofa-elliq>. Acesso em: 31 mai. de 2023.

WHITEMAN, Vivian. *Fome de outro: será que a dependência é tão ruim assim?* *Elle Brasil*. Disponível em: <https://elle.com.br/colunistas/depende-de-voce>. Acesso em: 16 mai. de 2023.



■ Giselle Beiguelman, frames do vídeo generativo da série Flora Rebellis, versão 2024.
Imagens geradas com Inteligência Artificial (Style Gans)

■ [Acesse o vídeo clicando na imagem](#)





A INTELIGÊNCIA ARTIFICIAL REPRESENTADA EM GENSHIN IMPACT,

iniciativas regulatórias e letramento algorítmico

DOI: <https://doi.org/10.35699/2965-6931.2023.47558>

MAURÍCIO FERREIRA SANTANA* GERALDO MAGELA PIERONI**

RESUMO: Uma das histórias do jogo Genshin Impact apresenta uma inteligência artificial, denominada Akasha, utilizada por cidadãos e acadêmicos da região de Sumeru. O protagonista da história, um viajante chamado Aether e sua companheira Paimon, chegam à região e passam a utilizar um aparelho que os conecta ao big data do Akasha; a história tem simbolicamente uma gradação que passa tanto pela tecnofilia quanto pelo tecnodeterminismo (Bernard Miège). É com base neste caráter simbólico que, lançando mão de análise de conteúdo – os diálogos da missão – transpomos esses elementos à realidade, buscando pontuar as iniciativas regulatórias governamentais sobre inteligência artificial (União Europeia, Estados Unidos, Brasil), bem como não governamentais (Future of Life Institute). Por fim, lançamos um olhar sobre as abordagens das ciências da comunicação a respeito do assunto (Lucia Santaella, David Gunkel et al.) e concluímos que são misteres a transdisciplinaridade e o letramento algorítmico.

PALAVRAS-CHAVE: comunicação; inteligência artificial; Genshin Impact; letramento; algoritmos.

Artificial intelligence represented in Genshin Impact, regulatory initiatives and algorithmic literacy

ABSTRACT: One of the stories in the game Genshin Impact features an artificial intelligence, called Akasha, used by citizens and academics in the Sumeru region. The protagonist of the story, a traveler named Aether and his partner Paimon, arrive in the region and start using a device that connects them to the big data of Akasha; the story symbolically has a gradation that passes through both technophilia and technodeterminism (Bernard Miège). It is based on this symbolic character that, using exploratory research with an emphasis on content analysis – the mission dialogues – we transpose these elements to reality, seeking to punctuate governmental regulatory initiatives on artificial intelligence (European Union, United States, Brazil), as well as non-governmental ones (Future of Life Institute). Finally, we look at the communication science approaches to the subject (Lucia Santaella, David Gunkel et al.) and conclude that transdisciplinarity and algorithmic literacy are essential.

KEYWORDS: communication; artificial intelligence; Genshin Impact; literacy; algorithms.

* Universidade Tuiuti do Paraná / ** Universidade Tuiuti do Paraná

Introdução

Nossa primeira reflexão ao escrever esse artigo remete às nossas escolhas: os autores que utilizamos como arcabouço teórico, a linha de raciocínio adotada para estabelecer um grau de entendimento aos leitores, o material empírico escolhido. Perguntamos se nossas escolhas são realmente nossas; ao utilizar os recursos de buscas bibliográficas, como as bases de dados acadêmicas e o Google Scholar, questionamos se os resultados são os que realmente queremos ou o que os algoritmos determinam, seja através de um ranqueamento pelo número de citações de determinado artigo, ou pela frequência de acessos de acordo com as palavras-chave que colocamos. O uso dos recursos de truncagem, a aplicação de operadores booleanos (*and*, *or*, *not*) são efetivos para os resultados que queremos? Se lançamos mão de alertas sobre determinados assuntos, recebemos o que há de mais relevante? São questionamentos que devem vir à consciência do pesquisador a todo momento, pois corremos o risco de ligar o “piloto automático” sem avaliar efetivamente o conhecimento que utilizamos. Observamos esse automatismo nas crianças e jovens (e aqui incluímos nossos filhos), quando “dar um Google” significa perguntar a uma espécie de oráculo qualquer assunto e receber a resposta como definitiva, verdadeira. Somos corresponsáveis na alimentação de *fake news* e na escassez de pensamento crítico. Sabemos muito bem as consequências disso. Acreditar em *fake news* e compartilhá-las sem nenhum tipo de verificação corrói a democracia e abre terreno para o fascismo e o extremismo – em nosso caso particular, citamos os acontecimentos recentes ocorridos no Brasil a partir de 2014 que culminaram no *impeachment* da presidente Dilma Rousseff e que levaram Jair Bolsonaro ao poder. Outros exemplos no mundo aludem à ascensão de Donald Trump ao poder nos Estados Unidos, à saída do Reino Unido da União Europeia, entre outros fatos.

Não temos como objetivo apresentar um quadro distópico que criminalize a inteligência artificial, que a condene por todos os males da sociedade moderna. O que nos motivou a escrever esse texto foi uma situação lúdica ao jogarmos o RPG (*Role-Playing Game*) *Genshin Impact*, dentro de uma história específica, a qual nos remeteu ao contexto descrito inicialmente, onde (aludindo a Aristóteles) “a arte imita a vida” ou pelo menos parece dar indícios de que a arte “alerta” a vida. Desenvolvemos esse raciocínio

adiante, pois temos como objetivo tensionar o que é mostrado no game sob o aspecto da Inteligência Artificial (IA) e seus desdobramentos, bem como despertar reflexões sobre como a IA está sendo vista hoje, tomando como ponto de partida o manifesto lançado em 2023 pelo Future of Life Institute. Por fim, apontamos que as teorias da comunicação precisam estabelecer chaves de entendimento para a IA não só como mídia, mas nos seus processos de regulação.

Alguns marcos da inteligência artificial

Não temos como objetivo incluir aqui um apanhado histórico da IA, salvo alguns marcos importantes. Vale mencionar que a obra de Stuart Russel e Peter Norvig destinada a cursos de Ciência da Computação oferece diversas abordagens conceituais sobre IA, dentre as quais destacam-se as voltadas para

processos de pensamento e raciocínio (...), comportamento (...), fidelidade ao desempenho humano e um conceito ideal de inteligência, chamado de racionalidade. Um sistema é racional se “faz a coisa certa”, dado o que ele sabe. (...) As quatro estratégias para o estudo da IA têm sido seguidas (...) por pessoas diferentes com métodos diferentes. Uma abordagem centrada nos seres humanos deve ser em parte uma ciência empírica, envolvendo hipóteses e confirmação experimental. Uma abordagem racionalista envolve uma combinação de matemática e engenharia. (RUSSEL; NORVIG, 2013, p. 24-25).

Encontrar um conceito único para IA é, portanto, um problema. Dentre os vários que encontramos, o conceito abaixo parece um resumo adequado:

Inteligência artificial (IA) (...) se propõe a desenvolver sistemas que simulem a capacidade humana na percepção de um problema, identificando seus componentes e, com isso, resolver problemas e propor/tomar decisões. Outra definição de Inteligência Artificial [é] a criação de sistemas inteligentes de computação capazes de realizar tarefas sem receber instruções diretas de humanos (os “robôs” são exemplos disso). (LOBO, 2018, p. 4).

Destacamos três marcos históricos que ajudaram a consolidar a IA. O primeiro deles é o experimento de Turing, elaborado em 1950. Nesse experimento, Alan Turing dispõe de três elementos (A, B e C), sendo um deles um homem (A), outro mulher (B), outro assumindo o papel de interrogador (C). Um dos objetivos desse experimento é

¹ Do verbo *chat* (bater papo, prostrar, conversar) e da abreviação de *robot* (robô); ambas as definições obtidas no Cambridge Dictionary, disponível em <https://dictionary.cambridge.org>

fazer com que o interrogador adivinhe corretamente o sexo dos interrogados. David J. Gunkel (2017) resume o experimento de Turing nos seguintes passos: o interrogador faz para A e B várias perguntas e, baseado nas respostas, tenta discernir se o respondente é uma mulher ou um homem. Essas respostas deveriam ser datilografadas ou mesmo feitas por meio de um telegravador, configurando uma espécie de comunicação (chat) mediada por computador. Além disso, outras variáveis são levadas em conta, como a possibilidade do interrogado ser substituído por uma máquina a fim de “enganar” o interrogador etc. (Gunkel, 2017, p. 6-7).

O segundo marco refere-se à aplicação chamada ELIZA, desenvolvida em 1966 por Joseph Weizenbaum, uma “aplicação simples de processamento de linguagem natural que era capaz de conversar com interrogadores humanos de maneira a parecer outra pessoa [e considerada] o primeiro chatterbot”. (Ibidem, p. 8). Apesar do sucesso do experimento, o próprio Weizenbaum, em estudo posterior de 1976, foi autocrítico em afirmar que a IA poderia representar ameaças à sociedade, sobretudo na iminência de que os seres humanos se tornassem algo como autômatos, perdendo sua autonomia e sua própria humanidade (Russel; Norvig, 2013, p. 1189).

Por fim, mas não menos instigante, temos o experimento realizado por John Searle durante os anos 1980 conhecido como quarto chinês (ou sala chinesa):

O sistema consiste em um ser humano, que compreende apenas o idioma inglês, equipado com um livro de regras escrito em inglês e diversas pilhas de papel, sendo algumas em branco e algumas com inscrições indecifráveis (portanto, o ser humano faz o papel da CPU, o livro de regras é o programa, e as pilhas de papel são o dispositivo de armazenamento). O sistema está em um quarto com uma pequena abertura para o exterior. Por essa abertura passam folhas de papel com símbolos indecifráveis [os testes foram também realizados com ideogramas chineses]. O ser humano encontra símbolos correspondentes no livro de regras e segue as instruções. As instruções podem incluir escrever símbolos em novas folhas de papel, encontrar símbolos nas pilhas, reorganizar as pilhas, e assim por diante. Eventualmente, as instruções farão com que um ou mais símbolos sejam transcritos em uma folha de papel que será repassada de volta ao mundo exterior. Então, Searle argumenta da seguinte forma: a pessoa no quarto não compreende chinês (dado). O livro de regras e as pilhas de papel, sendo apenas folhas de papel, não entendem chinês. Então, não está acontecendo nenhuma

compreensão de chinês. Por conseguinte, de acordo com Searle, a execução do programa correto não gera necessariamente a compreensão. (Ibidem, p. 1186).

O experimento de Searle, a princípio, teve como objetivo refutar o de Turing. Apesar da interessante discussão gerada pelos resultados obtidos com a sala chinesa – que, em resumo, demonstrou que um programa de computador não pensa como um ser humano, apenas reproduz padrões –, tal experimento não reduz o de Turing, pois este ainda é bastante utilizado como base para o embate tanto filosófico como científico acerca da IA, sobretudo se assumimos que as máquinas podem “aprender”, vide o campo do *machine learning* e *deep learning*, este último uma forma de aprendizagem por máquina que lança mão de *big data*, onde

sistemas como o Facebook aprendem, por meio dos mais variados algoritmos, o que os usuários gostam, quais assuntos são mais relevantes para eles, e, desta forma, alteram diariamente qual informação e como ela é entregue aos usuários em suas *timelines*. A criação de assistentes virtuais como Siri (Apple), Alexa (Amazon), Cortana (Microsoft), dentre outros, vem estreitando as relações homem/máquina e alterando a forma como nos relaciona[mos] com a tecnologia. (CARDOZO; FERRARI; BOARINI, 2020, p. 55).

Os exemplos e experimentos que citamos acima referem-se à chamada “IA fraca”, que nada mais é do que programas de computador criados para agir com certa inteligência que reproduza a humana, justamente com a finalidade de interagir com o ser humano e a ele dar informações solicitadas. Aos pensadores, segundo Russel e Norvig, fica a preocupação com a “IA forte”, cujos questionamentos são: “é possível que as máquinas ajam com inteligência, de modo semelhante às pessoas (...), elas realmente terão mentes conscientes? Quais são as implicações éticas de máquinas inteligentes?” (2013, p. 1173). São estas preocupações (não futuras, mas presentes) que em muitas situações retratam a IA forte dentro das ficções distópicas, sejam elas representadas em filmes de ficção científica, romances (ao estilo *cyberpunk*) e em *games*.

O *akasha pulsa*

2 Jogos multiplataforma ou crossplay são desenvolvidos para rodar em celulares, PC, tablets e consoles. Dessa forma, na modalidade multiplayer, um jogador que tem o jogo instalado em seu smartphone pode jogar com outro que o tenha em um console de videogame, por exemplo.

A desenvolvedora Hoyoverse (anteriormente denominada MiHoYo) foi fundada em 2012 por três estudantes universitários de Xangai (Liu Wei, Cai Haoyu and Luo Yuhao) e seus negócios têm foco em desenvolvimentos de jogos para celulares (*mobile games*), multiplataforma² e animações (Stoller et al., 2022, p. 36). Entre seus principais lançamentos, podemos citar *Guns GirlZ* (*mobile*, 2014), *Honkai Impact 3rd* (*mobile*, PC, 2016), *Genshin Impact* (multiplataforma, lançado na China em 2019 e nos demais países em 2020) e *Honkai: Star Rail* (multiplataforma, 2023).

Genshin Impact (GI) é um *Role-Playing Game* (RPG) *free-to-play* de mundo aberto e pode ser jogado em *single* ou, em algumas situações, *multiplayer*. O jogo teve o maior investimento até então da desenvolvedora, o qual foi motivado pela boa recepção do produto anterior, *Honkai Impact 3rd*:

[A chamada à época] MiHoYo levou 3 anos e aproximadamente 100 milhões de dólares de investimento neste projeto antes do jogo ser lançado em 2019 (...). Combinado com a mais recente tecnologia de modelagem 3D e recursos de jogo multiplayer, tornou-se um dos jogos mais populares na China e na indústria mundial de jogos móveis. (STOLLER et al., 2022, p. 37, tradução nossa).

A excelente recepção do público pode ser vista na quantidade de *downloads* efetuados. Em consulta somente ao *app* Google Play, detectamos que o jogo teve mais de 50 milhões de *downloads*. *Genshin Impact* também ganhou prêmios importantes no mundo *gamer* (dois Game Awards, respectivamente em 2021 e 2022).

A história do jogo se passa no mundo de Teyvat e é protagonizada por dois irmãos aventureiros que viajam por diversos mundos, Aether (personagem jogável da versão masculina) e Lumine (na versão feminina). No início da história, os irmãos viajam para Teyvat e um deles (Aether ou Lumine, dependendo da escolha do jogador) é sequestrado por uma personagem misteriosa. A jornada começa com a procura de um irmão por outro e se desdobra em várias histórias sequenciadas principais (denominadas “atos” ou “missão do arconte”), intercaladas por histórias paralelas (“missões de mundo”). O (a) protagonista tem como companheira Paimon, uma fada que auxilia nas missões, apesar de não ser um personagem jogável.



Figura 1 – Aether (direita) e Paimon (esquerda).

Fonte: Screenshot (smartphone pessoal)

Teyvat é dividida em sete nações (até o momento em que escrevemos este texto, somente quatro são conhecidas), cada qual com sua divindade (arcontes), possuidora de um elemento específico – são seis elementos constantes no jogo: anemo (vento); pyro (fogo); cryo (gelo), hydro (água), electro (raios), geo (terra e rochas) e dendro (sementes explosivas). Essas nações aludem a partes reais de nosso mundo: Mondstadt retrata países como França e Alemanha; Liyue representa a China; Inazuma, o Japão e Sumeru, um território misto que compõe tanto a Índia quanto partes da África e do Oriente Médio, como o Egito e o Irã. A divindade de Sumeru é conhecida por Nahida.



Figura 2 – Mosaico em sentido horário: Mondstadt, Liyue, Inazuma e Sumeru.

3 O gacha (ou gashapon) é um sistema que dá recompensas aleatórias a um jogador (originalmente lançado no Japão como uma máquina que, ao inserir uma moeda ou cédula, dava um brinde aleatório), podendo ser pago (em dinheiro real) ou gratuito (através da obtenção de moedas virtuais dadas como recompensas em missões). Para obter personagens ou itens valiosos, o jogador necessita dispendar grandes quantidades de dinheiro (virtual ou real). Não é nossa intenção discorrer sobre os problemas decorrentes desse sistema e de eventuais prejuízos financeiros ou psicológicos para jogadores, sobretudo crianças. Em relação a isso, no caso japonês, Fujihara e Shibuya afirmam que “(...) os gastos com gacha aumentaram rapidamente [a partir de 2003] e se tornaram um problema social no Japão.” (2020, p. 1, tradução nossa).

Além de Aether/Lumine, todos os arcontes são personagens jogáveis, bem como outros que vão surgindo à medida que o jogador avança na história, podendo ser obtidos através de banners periódicos disputados em gacha³.

Para esse estudo, lançamos mão do capítulo III da missão de Sumeru, dividida em 5 atos, dos quais extraímos recortes dos seguintes: “Através da névoa e da floresta” (Ato I), “O Alvorecer das Rosas” (Ato II), e o ato final (V) “O Akasha pulsa, a chama Kalpa se Ergue”. O gameplay dos cinco atos totaliza aproximadamente 18 horas de jogo (jogamos em smartphone). Todos os recortes que citamos adiante estão registrados no jogo textualmente, podendo ser acessados no menu “arquivo / diário de viagem / Sumeru”, com a tradução para o português brasileiro.

A história tem vários desdobramentos e grande número de personagens (como é de se esperar de um jogo de RPG). Avançamos 4 subcapítulos do primeiro ato da história para chegar ao ponto em que Aether (jogamos com o personagem masculino) e Paimon localizam a cidade de Sumeru em busca de Nahida, conhecida como deusa do conhecimento e da sabedoria. Eles supõem que, por ser detentora de tais qualificações, ela possa dar pistas sobre o desaparecimento de Lumine. No subcapítulo do Ato I “Em busca da deusa da sabedoria”, chegando aos portões da cidade, ambos encontram o guarda Panah. Este se dirige à dupla e fornece a eles um dispositivo, o terminal Akasha (o Akasha é uma tecnologia criada pela Arconte anterior à Nahida, chamada de Lorde Maior Rukkhadevata), o qual pode ser utilizado para busca de informações. Os dois, no entanto, ao consultar o dispositivo para buscar o paradeiro de Nahida, não encontram nenhuma informação.



Figura 3 – Aether utilizando o terminal Akasha em seu ouvido.

Fonte: Screenshot (smartphone pessoal).

No Ato II, o protagonista conhece Alhaitham, escriba da Academia de Sumeru que está investigando o contrabando de “conhecimento enlatado”, pequenas cápsulas (como *pen drives*) desenvolvidas pela Academia de Sumeru contendo tipos específicos de conhecimento, como matemática, astrologia etc. que foram extraídos da árvore do conhecimento Irminsul e podem ser acessadas pelo terminal Akasha. Neste mesmo ato Nahida se manifesta a Aether possuindo o corpo de outra pessoa; esta é a maneira com a qual ela pode se comunicar, visto que os habitantes de Sumeru estão presos a um samsara produzido por alguns dos sábios da Academia que visam roubar, através dos terminais Akasha, o sonho das pessoas para alimentar a *base de dados* do Akasha. Após descobrirem uma maneira de quebrar o ciclo do samsara, Aether e Nahida sofrem, no Ato III, uma emboscada dos sábios. Aether consegue escapar, mas Nahida é capturada.



Figura 4 – Nahida.

Fonte: Screenshot (smartphone pessoal).

O desfecho da história se dá no Ato V, quando Nahida consegue conectar sua consciência ao Akasha, para comunicar a Aether que está presa. Após várias batalhas, Aether, por meio de um terminal Akasha *hackeado* por Alhaitham, consegue entrar na consciência de Nahida e despertá-la para a liberdade. Nahida pune os sábios responsáveis pelos crimes cometidos e estabelece uma nova ordem em Sumeru, acabando com o Akasha e indicando um novo conselho de sábios de sua confiança para a Academia.

Feito este breve resumo da história, podemos nos debruçar um pouco sobre o que é o Akasha e o que ele representa para o povo de Sumeru. A nós fica claro, ao

estabelecer um parâmetro com a realidade, que esta tecnologia é poderosa o bastante para ser considerada como uma IA, tanto que ao ser inserida no ouvido do usuário, ela parece identificar automaticamente qual seu comportamento via ondas cerebrais, a fim de fornecer as informações mais precisas. Um dos elementos que compõe esta arquitetura tecnológica é a Irminsul, cuja representação é uma gigantesca árvore que retêm todos os registros da história de Sumeru, bem como a história individual de cada cidadão, vivo ou falecido. Na mitologia nórdica, Irminsul é a árvore centro do mundo, que possui o saber e o conhecimento e era representada por um pilar (Langer, 2007, p. 66). Além disso, da Irminsul se extrai o “conhecimento enlatado”, ou “cápsulas do conhecimento”, de uso estritamente acadêmico e que, de maneira ampla, é o banco de dados fonte para a IA Akasha, que pode ser consultada pelo *motor de busca* terminal Akasha. Abaixo, representamos como esses elementos formam um sistema:



Figura 5 – Arquitetura da IA Akasha. Fonte: Nossa autoria.

Lembrando que muitos elementos em Sumeru são referências à cultura indiana, Akasha, em sânscrito, significa

éter, espaço que permeia tudo, considerado na filosofia indiana o primeiro e o mais fundamental dos cinco elementos – vata (ar), agni (fogo), ap (água) e prithivi (terra). Akasha reúne as propriedades dos cinco elementos e é o ventre de onde emergiu tudo o que percebemos com nossos sentidos e para dentro de onde tudo no final deverá imergir. Akasha é o meio de onde tudo emergiu, emerge e emergirá e para o qual tudo retorna, não sem antes deixar sua memória nele gravada. (SOUSSUMI, 2018, p. 212, grifo nosso).

A expressão Akasha foi adequadamente atribuída no jogo, pois, segundo a arquitetura que apresentamos, se na cultura indiana ele é um dos cinco elementos onde tudo emerge e para o qual tudo retorna, nos remete diretamente à teoria dos sistemas – “sistema é um conjunto de elementos que interagem para realizar objetivos.” (Stair; Reynolds, 2015, p. 8) – e à cibernética. Vide o conceito dado em 1948 por Norbert Wiener:

Uma ação complexa é aquela em que os dados introduzidos, (a que chamamos entrada) para obter um efeito sobre o mundo exterior — efeito a que chamamos saída — podem implicar um grande número de combinações. Combinações dos dados introduzidos no momento com os registros obtidos de dados anteriores armazenados, a que chamamos memória, e que estão registrados na máquina [...]. Tal controle da máquina com base no seu desempenho efetivo em vez de no seu desempenho esperado é conhecido como realimentação (feedback). (WIENER, 1968, p. 23-24).

Na arquitetura que propusemos, temos então as entradas (*inputs*) de dados e informações (como os sonhos e as teses acadêmicas dos sábios), a unidade de processamento (Akasha) que acessa o grande banco de dados Irminsul para fornecer as saídas (*outputs*), respostas às requisições e cápsulas de conhecimento.

Selecionamos alguns diálogos da história para ilustrar a gradação narrativa em relação ao Akasha, que inicia pela fase do deslumbre e termina com o seu banimento. Do capítulo III do Ato I, trazemos estes diálogos:

Panah: Enquanto você estiver dentro das fronteiras de Sumeru, você pode usar o Terminal Akasha para se conectar diretamente com o Akasha e acessar qualquer conhecimento que você quiser (...).

Paimon: Uau! Alguma coisa aconteceu [após inserir o terminal Akasha em seu ouvido] e agora Paimon sabe exatamente como usar essa coisa! Parece que tudo o que você precisa fazer é se concentrar naquilo que você quer saber e... BAM! Você recebe o conhecimento! Nossa, isso é realmente muito útil! (...)

Aether: Além disso, mesmo que a mesma pergunta seja feita por várias pessoas, o Akasha ainda assim transmitiria o conhecimento baseado na idade, identidade, experiência e outros traços de quem está perguntando.

Esse deslumbre de Paimon com uma tecnologia que até então não conhecia remete diretamente ao entusiasmo do mundo *online*, sobretudo nos anos 2000, onde autores como Jean Burgess e Joshua Green exaltaram o YouTube (2009, p. 23), Pierre Levy, em seus primeiros escritos, a chamada “inteligência coletiva” (2003, p. 28) e

Henry Jenkins com a “cultura participativa” (2009, p. 341). A devida cautela foi dada no amadurecimento dos estudos dos pesquisadores, mas esse encantamento pertence também ao usuário comum de redes sociais e motores de busca como o Google. A facilidade do recurso obviamente leva à comodidade do usuário, o qual não precisa elaborar estratégias de busca (a exemplo das bases de dados científicas), utilizando quaisquer termos soltos – inclusive por comando de voz – que trazem respostas correspondentes. Quando se tem acesso ao perfil do usuário por meio da permissão dada pelo mesmo – onde em muitos casos mal se lê os contratos de privacidade e permissão de uso – a IA, via *machine learning*, fornece, conforme a fala de Aether, informações específicas de acordo com as características e perfis de busca registrados. No *game*, isso é representado pelos sonhos dos habitantes de Sumeru “contrabandeados” para o Akasha, conforme extratos do capítulo III, ato II:

Paimon [perguntando à Nahida]: as pessoas em Sumeru acham que não sonham, mas a verdade é que o Akasha rouba seus sonhos sem elas saberem, correto?

Nahida: sonhos são fantásticos, complexos e cheios de imaginação. O cérebro das pessoas está mais ativo quando elas estão dormindo. Noutras palavras, sonhos são ricos compêndios de sabedoria humana.

Ou seja, o roubo dos sonhos sem sabê-lo equivale à captação dos dados pessoais e de preferências de usuários por meios ilegais (acessos indevidos aos perfis) ou, conforme comentamos, por meios legais, via contrato de privacidade e acesso a dados pessoais que não são devidamente lidos – clica-se em “aceitar” para utilização imediata de uma rede social, por exemplo. É nesse momento que surgem movimentos criminosos os mais diversos, que vão desde a extorsão e chantagem pessoais como utilização em massa de milhões de perfis visando alterações no curso da política e democracia, onde citamos a Cambridge Analytica, nos casos *Brexit*, no Reino Unido (concretizado em 2020) e na eleição de Donald Trump, em 2016 (o mesmo *modus operandi* utilizado nas eleições brasileiras de 2019), onde a principal matéria-prima foram os perfis do Facebook, através da “prática de ataques-focais (*microtargeting*, em inglês) de seus usuários, muitas vezes utilizando-se – de forma intencional – de notícias falsas (*fake News*) para manipular tendências políticas de eleitores” (Fornasier; Beck, 2020, p. 184). O problema das *fake News* também não passa despercebidamente em *Genshin Impact*, vide os diálogos do capítulo III, Ato V, abaixo:

Alhaitham: farei uma cápsula de conhecimento idêntica, mas dessa vez, colocarei algumas informações mentirosas nela [para alimentar o Akasha].

Dehya: tenho apenas uma pergunta. Mesmo que a informação falsa possa ser colocada na cabeça dos guardas, como poderemos ter certeza de que eles vão acreditar?

Alhaitham: as pessoas que estão acostumadas a confiar no Akasha se tornam menos inquisitivas, e sua habilidade de distinguir a verdade da desinformação é reduzida significativamente... Eles veem tudo que o Akasha diz como a verdade incontestável.

Cyno: o Akasha irá transformar os humanos em “máquinas”.

Apesar do *hacking* de Alhaitham ter o propósito de ajudar Aether e Nahida, “plantando” mentiras, na sequência do diálogo a personagem Dehya questiona sobre a efetividade das informações manipuladas, quando é prontamente respondida por Alhaitham e o personagem Cyno, os quais afirmam que os usuários do Akasha já não conseguem mais diferenciar o que é uma informação real de uma falsa, pois confiam cegamente no sistema (o Akasha tem a verdade incontestável).

A atitude criminosa dos sábios da Academia e o uso indevido do Akasha faz com que Nahida opte por desativá-lo. Ainda no capítulo III, Ato V:

Nahida: O futuro da cidade de Sumeru estará apenas nas minhas mãos. Eu vou desativar o Akasha e deixar que a curiosidade e a sede por conhecimento propulsionem o reino acadêmico mais uma vez...

Nahida: Após os eventos recentes, o Akasha não pode mais funcionar como antes / Eu pensei um pouco e decidi desativá-lo permanentemente / Isso definitivamente não é algo ruim. (...) / A administração centralizada do conhecimento pelo Akasha sempre restringiu a curiosidade das pessoas e reduziu o número de caminhos disponíveis a elas. Eu não acho que isso seja bom para Sumeru.

A atitude peremptória e aparentemente despótica de Nahida radicalmente condena a centralidade do Akasha como única fonte de conhecimento; sua decisão em desativá-lo visa a reeducação dos cidadãos e acadêmicos para o raciocínio e a pesquisa: em resumo, Nahida busca a promoção de um novo tipo de letramento que não o digital – mesmo porque tal letramento nunca existiu. No jogo, portanto, a opção mais fácil foi a extinção da tecnologia sob forma de *downgrade* sociotécnico-cultural para a retomada cognitiva da leitura e da escrita. No mundo real, extinguir a IA está totalmente fora de questão, vide a predominância econômica das *big techs* e o enraizamento

4 Segundo o website do FOLI, sua missão é “direcionar a tecnologia transformadora para beneficiar a vida e longe de riscos extremos em grande escala.” (FOLI, 2023). Um de seus patrocinadores é o bilionário Elon Musk (SpaceX, Tesla, Twitter); além da IA, as outras áreas abrangidas pelo Instituto são biotecnologia, armamento nuclear e mudanças climáticas.

das tecnologias de *machine learning* sobre inúmeros campos da ciência. Somos então levados a coexistir com essas tecnologias, mesmo que a maioria de seus usuários não tenha o letramento adequado para utilizá-la de maneira consciente e mesmo crítica, o que nos deixa ignorantes dos avanços da IA e de seu potencial risco para a autonomia de pensamento do ser humano.

Parem a ia!

Tal preocupação vem ganhando espaço na voz de vários cientistas ao redor do mundo. Notadamente chamou a atenção dos meios de comunicação a carta aberta do Future Of Life Institute⁴ (Foli) publicada em 22 de março de 2023, denominada “Pause Giant AI Experiments: An Open Letter”, cujo subtítulo é “We call on all AI labs to immediately pause for at least 6 months the training of AI systems more powerful than GPT”; em tradução livre, “Pausar experimentos das gigantes de IA: uma carta aberta”, “convocamos todos os laboratórios de IA a imediatamente pausar por pelo menos 6 meses o treinamento de sistemas de IA mais poderosos que o GPT”, respectivamente. Essa carta aberta é baseada nos The Asilomar AI Principles definidos na Beneficial AI 2017, conferência promovida pelo próprio Foli – simbolicamente, no mesmo local em que ocorreu a conferência para regulação da pesquisa em manipulação genética, em 1975 – e que, em linhas gerais, propõe princípios de governança para a IA, dentre os quais nas esferas da pesquisa, ética e valores, e questões de longo prazo, como a superinteligência (artificial) (Foli, 2023; todas as traduções daqui em diante são nossas).

Dentre os pontos apresentados da carta, que possuem referencial teórico – a bibliografia utilizada consta nas notas finais – destacamos alguns: o texto inicia com uma reflexão geral, de que “a IA avançada pode representar uma mudança profunda na história da vida na Terra, e devem ser planejados e administrados proporcionalmente com cuidados e recursos” (Ibidem), para em seguida lançar algumas questões que necessitam de respostas:

Os sistemas contemporâneos de IA estão competindo com humanos em tarefas gerais (...); devemos deixar que as máquinas inundem nossos canais de informação com propaganda e falsidade (...), automatizar todos os trabalhos (...), desenvolver não-humanos? (...). Devemos arriscar a perda do controle de nossa civilização? Tais deci-

sões não devem ser delegadas a líderes tecnológicos não eleitos. Sistemas poderosos de IA devem ser desenvolvidos apenas quando estivermos confiantes de que seus efeitos serão positivos e seus riscos, administráveis. (Ibidem).

A carta afirma, ainda, que vivemos uma espécie de “verão de IA”, ou seja, um momento extremamente profícuo no desenvolvimento e implantação dessas tecnologias, mas que é urgente que existam diretrizes de governança para regulação, e que esta regulação deva vir dos países:

Os desenvolvedores de IA devem trabalhar com os formuladores de políticas [governamentais] para acelerar (...) o desenvolvimento de sistemas robustos de governança de IA. Estes devem, no mínimo, incluir (...) autoridades reguladoras dedicadas à IA (...); um ecossistema robusto de auditoria e certificação; responsabilidade por danos; financiamento público robusto para pesquisa técnica de segurança de IA; e (...) recursos para lidar com as dramáticas perturbações econômicas e políticas (especialmente para a democracia) que a IA poderá causar. (FOLI, 2023).

No caso específico da desenvolvedora OpenAI, citada no documento por conta do desenvolvimento do ChatGPT, o documento afirma que esta empresa, por visão auto-crítica, afirma que “em algum momento, pode ser importante obter revisão independente antes de começar a treinar sistemas futuros e para os esforços mais avançados concordarem em limitar a taxa de crescimento da computação usada para criar novos modelos.” (FOLI, 2023). Mas, quem é a OpenAI e o que é o ChatGPT?

Segundo informação obtida no site da empresa, “a OpenAI é uma empresa de pesquisa e desenvolvimento de IA. Nossa missão é garantir que a inteligência artificial geral⁵ beneficie toda a humanidade.” (Openai, 2023, tradução nossa). Entre seus principais produtos de IA estão o ChatGPT e o DALL-E. A tecnologia GPT (Generate Pre-trained Transformer, ou transformador pré-treinado generativo) utiliza machine learning e funciona como chatterbot, mas apenas com inputs e outputs textuais na versão atual (o GPT4 está sendo aprimorado para inputs de imagens), enquanto o DALL-E funciona com inputs textuais para a geração de outputs tanto textuais quanto visuais, ou seja, pode-se solicitar uma ilustração de qualquer tipo que o DALL-E criará.

Diferentemente da personagem Nahida, o manifesto do Foli não propõe a extinção da IA, mas, baseados no avanço promissor dessa tecnologia, representada pela liderança da OpenAI, propõe que os governos dos países sejam mais atuantes para a elabora-

5 A inteligência artificial geral existe quando o sistema tem um processador adequadamente programado, uma ‘mente’, com entradas e saídas corretas, no sentido em que os humanos têm mentes. É um sistema com capacidade de aplicar inteligência a qualquer problema e não só a uma tarefa ou problema específico. (Parreira; Lehmann; Oliveira, 2021, p. 980).

ção de diretrizes de governança que possibilitem a devida regulação da IA, visando a segurança tanto dos próprios sistemas – ou seja, a IA deve operar em níveis seguros que possam a qualquer momento viabilizar a interferência humana – como do próprio ser humano sob os aspectos éticos e morais.

Em relação às ações governamentais, no primeiro semestre de 2019 a OCDE (Organização para a Cooperação e Desenvolvimento Econômico) anunciou princípios para o desenvolvimento de inteligência artificial, sendo o Brasil um dos signatários do documento, os quais totalizam 42 países. O documento da OCDE recomenda que os aderentes promovam e implementem os “princípios éticos para a administração responsável de IA” (Brasil, 2020, p. 7).

A União Europeia (UE), por sua vez, elaborou o “The Artificial Intelligence Act” (Bruxelas, 2021), o qual classifica as IAs de acordo com seu grau de risco: inaceitável, elevado e baixo ou mínimo (Comissão Europeia, 2021); o ato regulador foi aprovado em 14 de junho de 2023. Nos Estados Unidos, o Senado está em fase de audiências sobre o assunto. Segundo o senador Chuck Schumer, é necessário “elaborar uma legislação de IA que incentive a inovação, mas tenha as salvaguardas para evitar os riscos que a IA pode apresentar” (Bsartz, 2023).

No caso brasileiro, foi redigido o Projeto de Lei (PL) n. 2338/2023¹. Seu artigo primeiro estabelece

normas gerais de caráter nacional para o desenvolvimento, implementação e uso responsável de sistemas de inteligência artificial (...) no Brasil, com o objetivo de proteger os direitos fundamentais e garantir a implementação de sistemas seguros e confiáveis, em benefício da pessoa humana, do regime democrático e do desenvolvimento científico e tecnológico (BRASIL, 2023, p. 2).

O texto do PL também categoriza as IAs sob duas formas: risco excessivo e alto risco; apresenta também pontos que de certa forma convergem com as proposições do FOLI, como respeito aos direitos humanos e aos valores democráticos; o uso da IA para buscar resultados benéficos para as pessoas e o planeta, a garantia de transparência sobre o uso e funcionamento dos sistemas de IA e de divulgação responsável do conhecimento de IA, observados os segredos comercial e industrial, e de conscientização das partes interessadas sobre suas interações com os sistemas, inclusive no local

¹ A este PL estão pensados outros três: 5051/2019, 21/2020 e 872/2021.

de trabalho (Ibidem, p. 2-33); o PL no entanto minimiza a importância do letramento digital nesta tecnologia, mencionando apenas a “conscientização sobre os sistemas de inteligência artificial e suas aplicações”. (Ibidem, p. 3).

As ciências da comunicação e ia: rumo às considerações finais

Partindo da ideia de mediatização como uma interferência / influência (confluência, convergência etc.) das mídias nos processos sociais, apontando para possíveis mudanças nas formas e esferas das culturas, inferimos que uma interface de IA, como o *chatbot*, seja considerada uma mídia, porque evidentemente essa interface atua como mediador junto aos *scripts* algorítmicos para emitir informações, configurado grosso modo em um processo de comunicação mediatizado. No entanto, quando nos referimos à mediatização em larga escala, ou em âmbito de uma comunidade e mesmo globalmente aludimos a um

quadro sociológico que permite considerar como os meios de comunicação, em determinado período histórico da alta modernidade, transformaram-se em instituições semi-independentes na sociedade, ao mesmo tempo que se integraram a uma variedade de instituições sociais e culturais, e também como, por meio desse processo de diferenciação, integração e interação com outras instituições sociais e culturais, vieram e vêm a influenciar a sociedade. (Harvard, 2014, p. 243).

As considerações de Stig Hjarvard, dentro do que o autor chama de “lógicas da mídia”, parece apontar para um pensamento tecnodeterminista, mas não o é. Quando nos referimos ao tecnodeterminismo estamos apontando para uma perspectiva apocalíptica e irreversível de que nossos destinos estão nas mãos das tecnologias de informação e comunicação, e que elas ditam os rumos da humanidade. Tal pavor se potencializa quando o assunto é IA.

Para Bernard Miège (2009), a questão tecnodeterminista é em parte responsabilidade dos futurólogos e dos tecnófilos de plantão, que vislumbram um cenário na maior parte das vezes benéfico à humanidade e que não deixa de ser um contexto dúbio, pois favorece o pensamento de que cabe à tecnologia (corretamente denominadas

pelo autor como técnicas) – leia-se *big techs* – deixar fazer as coisas por si mesmas. Colocar o “carro na frente dos bois” é antecipar os usos de algo que ainda está por vir:

O exercício ao qual se entregam aqueles que regularmente se arriscam a prever o destino dos novos serviços merece raramente ser qualificado de prospectivo (...); as antecipações propostas não são objeto de verificações ulteriores e seus autores se premunem dos riscos de erro ao fazer listas nas quais o leitor está convidado a escolher e nas quais não deixamos de encontrar algumas verificações *ex post*, mas finalmente muito poucas. (MIÈGE, 2009, p. 28).

Trata-se de uma espécie de euforia sem conhecimento de causa, a *mídia* exaltando novas mídias, transformando-se “em transmissoras desse interesse indiscutível pela técnica, jornalistas e ensaístas reproduzindo, mais ou menos conscientemente, visões tecnicistas, atrativas e por vezes abertamente favoráveis [e que] quase não são postas em dúvida.” (Ibidem, p. 26). E é justamente essa falta de conhecimento que abre brechas para o tecnodeterminismo cego: para Miège, o desconhecimento bastante generalizado da esfera técnica e a redução desta a uma espécie de “caixa preta” favorece mais facilmente tanto a permanência de concepções tecnodeterministas quanto reprime e limita o debate público sobre as novas ciências e técnicas (Ibidem, p. 38).

Além da mediatização como parte dos estudos comunicacionais em IA, David J. Gunkel aponta que “as regras do jogo na teoria da comunicação precisam ser ajustadas e modificadas de forma significativa.” (2017, p. 15). O autor menciona a superficialidade dos estudos comunicacionais em tratar o computador como um meio de interação comunicativa, mas sem explorar o fato de que as máquinas podem ser *participantes* em interações comunicativas (2017, p. 16).

Não há efetivamente uma definição clara sobre quais aspectos da comunicação devam ser explorados para lidar com a IA e seus desdobramentos. Lucia Santaella tem como hipótese que “a comunicação como área de conhecimento está cada vez mais tomando o lugar de uma ciência piloto para cujas questões acabam convergindo muitas outras ciências” (2001, p. 4) resultando em redes de comunicação que têm interfaces e cooperam com outras disciplinas, incluindo a IA (Ibidem, p. 5). Em estudo mais recente, Santaella compara a desigualdade nos usos da IA tanto no Norte quanto no Sul Global, e no caso específico da produção cultural, alega que os “agentes culturais não têm conhecimento técnico na linguagem de programação (...) para criar conteúdos

por meio de aprendizagem de máquina” (2023, p. 29). David J. Gunkel considera que “comunicação e inteligência artificial estão intimamente relacionadas. Por um lado, a comunicação vem sendo um instrumental tanto para a teoria quanto para a prática da inteligência artificial” (2017, p. 6); ele enxerga novos desafios e oportunidades para os estudos em comunicação, porém reconhece que deverá haver uma recompilação da disciplina “e de seu foco tradicionalmente antropocêntrico” (Ibidem, p. 6). Maria Ignes Carlos Magno e Richard Romancini inferem que “o sentido social dos produtos culturais se elabora a partir de usos e apropriações – em determinado contexto de mediações –, nem sempre previsíveis. Compreender e dar direcionamento ético e humanista às práticas e à pesquisa que envolve a IA exigirá reflexão e investigação” (2023, p. 4). Em relação aos usos, Missila Cardozo, Pollyana Ferrari e Margareth Boarini apontam que nenhuma “tecnologia em si pode ser culpabilizada por usos inadequados ou por consequências indesejáveis. Até porque a tecnologia é neutra, por princípio. É o uso que se faz dela que a polariza” (2020, p. 62).

Daniel Melo Ribeiro e Geane Alzamora justificam a importância dos estudos comunicacionais voltados para IA à luz da semiótica. Segundo os autores, as pesquisas em IA se aproximam da semiótica pois computadores são máquinas processadoras de signos (baseadas em tipos de linguagem). A chave de investigação é inferir se “o processamento computacional é capaz de criar sentidos interpretativos sobre o mundo, sobre a realidade e sobre o contexto de significação” (2023, p. 12). Outras abordagens também são possíveis, como na análise estética – IAs que criam imagens inéditas, produzidas via programação com milhares de bancos de imagens de artistas consagrados – e na questão da desinformação, preocupação cara à disciplina do jornalismo (Ibidem, p. 12-14).

É consenso entre os citados autores a urgência no estabelecimento de políticas de regulação (ou políticas midiáticas) no uso de IA, em particular o uso de *chatbots*, que também envolvam aspectos éticos e humanistas, sob uma perspectiva multidisciplinar. Em relação a isso, há convergência tanto nos pontos supracitados do FOLI, quanto nos aspectos legais europeus e brasileiros em tramitação (governança, letramento digital). Outro ponto observado é que o uso da mídia IA deve estar condicionado à responsabilidade social de forma construtiva e em constante diálogo (Santaella, 2001, p. 5; 2023, p. 38; Hjarvard, 2014, p. 246; Gunkel, 2017, p. 17; Cardozo; Ferrarej; Boarini, 2020, p. 62; Magno; Romancini, 2023, p. 4).

Voltamos então às reflexões iniciais deste texto: se nossas escolhas são efetivamente *nossas* escolhas... ou acabamos por nos deixar seduzir em uma espécie de *samsara* criado propositalmente pelos acadêmicos de Sumeru (ou Vale do Silício?) que nos mergulham em torpor, fazendo-nos recorrer aos oráculos Google, ChatGPT, DALL-E ou algum outro que surja enquanto aqui escrevemos. Este *samsara* de Sumeru, círculo vicioso e en(torpe)cente, onde vivemos o mesmo dia incontáveis vezes só pode ser rompido através do despertar de uma consciência que enfoque o correto uso da IA. Mas, qual é o padrão ideal de uso? Como podemos distinguir o “bom” uso do “mau” uso? E pior, poderíamos estar sendo usados para usar? É nesse sentido que o *stand by* proposto pelo FOLI é oportuno, para (re)pensarmos nossos papéis enquanto estudiosos da comunicação frente a essa “caixa preta” da IA, missão também destinada aos governos, os que sofreram ataques à democracia – e dolorosamente mencionamos os ataques golpistas por vândalos aos prédios dos três poderes brasileiros em 8 de janeiro de 2023 – e os que potencialmente possam sofrer ataques. Nesse sentido, recomendamos que os governos iniciem a queda de braço com as *big techs* para ter o direito de realizar auditorias algorítmicas por especialistas, visando evitar tanto *scripts* quanto algoritmos de caráter enviesado que promovam a desinformação.

As Ciências da Comunicação necessitam honrar o seu caráter trans e multidisciplinar, agregando novas disciplinas – mesmo que em um nível de entendimento teórico-filosófico – como a Ciência da Computação. Com isso, queremos dizer que devemos dispende algum esforço para entender como funciona a lógica algorítmica, a fim de desvelar a malfadada caixa preta da tecnologia; conhecer o *modus operandi* do pensamento informático, agregando com isso subsídios para o aprimoramento das teorias comunicacionais a fim de reivindicar não o protagonismo, mas um papel contribuinte na construção de ideias que auxiliem a compreensão das mídias presentes e o processo de mediatização, sem exercícios prospectivos. Não temos Nahida para nos salvar do pesadelo tecnodeterminista e justamente por isso precisamos de letramento algorítmico.

Referências

BARTZ, Diane. Senado dos EUA realizará audiências sobre o uso de inteligência artificial. CNN / Reuters, 23 jun. 2023. Disponível em: <https://www.cnnbrasil.com.br/tecnologia/senado-dos-eua-realizara-audiencias-sobre-o-uso-de-inteligencia-artificial/>. Acesso em: 2 ago. 2023.

BRASIL. Projeto de lei da Câmara dos Deputados n. 21, de 2020: estabelece princípios, direitos e deveres para o uso de inteligência artificial no Brasil, e dá outras providências. Brasília: Senado Federal, 2021. Disponível em: https://legis.senado.leg.br/sdleg-getter/documento?dm=9063365&ts=1689259287250&disposition=inline&_gl=1*1r3fu1t*_ga*NDk1MTk1OTE1LjE2ODQxODI2ODE.*_ga_CW3ZH25XMK*MTY5MDI4NzYzNi43LjEuMTY5MDI4NzcoNS4wLjAuMA.. Acesso em: 23 jul. 2023.

____. Projeto de lei n. 2338, de 2023: dispõe sobre o uso da Inteligência Artificial. Brasília: Senado Federal, 2023. Disponível em: https://legis.senado.leg.br/sdleg-getter/documento?dm=9347622&ts=1689259290825&disposition=inline&_gl=1*18xz1g8*_ga*NDk1MTk1OTE1LjE2ODQxODI2ODE.*_ga_CW3ZH25XMK*MTY5MDk3Njk4Mi44LjEuMTY5MDk3NzMyMC4wLjAuMA.. Acesso em: 2 ago. 2023.

BURGUESS, Jean.; GREEN, Joshua. YouTube e a revolução digital: como o maior fenômeno da cultura participativa transformou a mídia e a sociedade. São Paulo: Aleph, 2009.

CARDOZO, Missila; FERRARI, Pollyana; BOARINI, Margareth. A inteligência artificial reconfigura a dinâmica comunicacional. Paradoxos, Uberlândia, v. 5, n. 1, p. 49-65, jan./jun. 2020. DOI: <http://doi.org/10.14393/par-v5n1-2020-54430>. Acesso em: 10 jun. 2023.

COMISSÃO EUROPEIA. Proposta de Regulamento do Parlamento Europeu e do Conselho que estabelece regras harmonizadas em matéria de inteligência artificial (Regulamento Inteligência Artificial) e altera determinados atos legislativos da União. Bruxelas: Comissão Europeia. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:52021PC0206&from=EN>. Acesso em: 24 jul. 2023.

FORNASIER, Mateus de Oliveira; BECK, Cesar. Cambridge Analytica: escândalo, legado e possíveis futuros para a democracia. Revista Direito em Debate, v. 29, n. 53, p. 182-195, jan./jun. 2020. DOI: <https://doi.org/10.21527/2176-6622.2020.53.182-195>. Acesso em: 12 jul. 2023.

FUJIIHARA, Masahito; SHIBUYA, Akiko. How is the gacha system reported on in Japan? Proceedings of DiGRA, p. 1-5, jun. 2020. Disponível em: http://www.digra.org/wp-content/uploads/digital-library/DiGRA_2020_paper_89.pdf. Acesso em: 25 jul. 2023.

FUTURE OF LIFE INSTITUTE. Pause Giant AI Experiments: An Open Letter. [S. l.]: Future of Life Institute, 22 mar. 2023. Disponível em: <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>. Acesso em: 25 jul. 2023.

GUNKEL, David. Comunicação e inteligência artificial: novos desafios e oportunidades para a pesquisa em comunicação. Galáxia, São Paulo, n. 34, jan./abr. 2017, p. 05-19. DOI: [dx.doi.org/10.1590/1982-2554201730816](https://doi.org/10.1590/1982-2554201730816). Acesso em: 20 jul. 2023.

HJARVARD, Stig. A mediatização da cultura e da sociedade. São Leopoldo: Ed. Unisinos, 2014.

JENKINS, Henry. Cultura da convergência. São Paulo: Aleph, 2009.

LANGER, Johnni. O mito do dragão na Escandinávia parte dois: as Eddas e o sistema ragnarokia-

no. *Brathair – Revista de Estudos Celtas e Germânicos*, v. 7, n. 1, p. 59-95, 2007. Disponível em: <https://ppg.revistas.uema.br/index.php/brathair/article/view/550>. Acesso em: 20 jul. 2023.

LÉVY, Pierre. *A inteligência coletiva: por uma antropologia do ciberespaço*. São Paulo: Loyola, 2003.

LOBO, Luiz Carlos. *Inteligência artificial, o futuro da medicina e a educação médica*. *Revista Brasileira de Educação Médica*, v. 42, n. 3, p. 3-8, 2018. DOI: <http://dx.doi.org/10.1590/1981-52712015v42n3R-B20180115EDITORIAL1>. Acesso em: 2 jul. 2023.

MAGNO, Maria Ignes Carlos; ROMANCINI, Richard. *A inteligência artificial, a inteligência humana e a pesquisa em comunicação*. *MATRIZES*, v. 17, n. 1, p. 3-8, jan./abr. 2023. DOI: <http://dx.doi.org/10.11606/issn.1982-8160.v17i1p3-8>. Acesso em: 23 jul. 2023.

NORVIG, Peter; RUSSEL, Stuart Jonathan. *Inteligência artificial*. Rio de Janeiro: Elsevier, 2013.

PARREIRA, Artur; LEHMANN, Lúcia; OLIVEIRA, Mariana. *O desafio das tecnologias de inteligência artificial na educação: percepção e avaliação dos professores. Ensaio: avaliação e políticas públicas em educação*, Rio de Janeiro, v. 29, n. 113, p. 975-999, out./dez. 2021. DOI: <https://doi.org/10.1590/S0104-40362020002803115>. Acesso em: 25 jul. 2023.

SANTAELLA, Lucia. *Novos desafios da comunicação*. *Lumina*, Juiz de Fora, v. 4, n. 1, p. 1-10, jan./jun. 2001. Disponível em: <https://www.ufff.br/facom/files/2013/03/R5-Lucia.pdf>. Acesso em: 2 jun. 2023.

_____. *Pensar a inteligência artificial: cultura de plataforma e desafios à criatividade*. Belo Horizonte: UFMG, 2023. Disponível em: <https://seloppgcomufmg.com.br/publicacao/pensar-a-inteligencia-artificial-cultura-de-plataforma-e-desafios-a-criatividade/>. Acesso em: 20 jul. 2023.

SOUSSUMI, Yusaku. *Por uma psicanálise em tempos de multiverso*. *Revista Multiverso*, v. 1, n. 1, p. 206-217, 2018. Disponível em: <https://www.bivipsi.org/wp-content/uploads/NPAracaju-multiverso-2018-v-1-n1-20.pdf>. Acesso em: 3 jul. 2023.

STAIR, Ralph M.; REYNOLDS, George W. *Princípios de sistemas de informação*. São Paulo: Cengage Learning, 2015.

STOLLER, Gregory [et al.]. *miHoYo: the open-world journey of an Otaku*. *Academy of Asian Business Review*, v. 8, n. 2, p. 33-60, 2022. DOI: <https://doi.org/10.26816/aabr.8.2.202212.33>. Acesso em: 1 jun. 2023.

WIENER, Norbert. *Cibernética e sociedade: o uso humano de seres humanos*. São Paulo: Cultrix, 1968.



■ Giselle Beiguelman, frames do vídeo generativo da série Flora Rebellis, versão 2024.
Imagens geradas com Inteligência Artificial (Style Gans)

■ [Acesse o vídeo clicando na imagem](#)





SUPERFÍCIE ARTIFICIAL, INTERIOR VISCERAL:

Aspectos tecnológicos, estéticos e econômicos da arte generativa

DOI: <https://doi.org/10.35699/2965-6931.2023.47661>

RAFAEL GALVÃO DE ALMEIDA*

RESUMO: A disseminação da arte gerada por inteligência artificial, ou arte generativa, está mudando as indústrias criativas. Este artigo discute questões tecnológicas, estéticas e econômicas da arte generativa. Apesar da tecnologia possibilitar novas formas de expressão artística, o status da arte generativa como arte não é certo. Defensores a veem como uma forma de arte não antropocêntrica, críticos enfatizam que a arte é feita por pessoas. A economia da cultura pode entender aspectos econômicos da arte generativa, entre elas a decisão de a adotar em larga escala para diminuir o efeito da doença dos custos; ela também demonstra que existe uma cadeia de trabalhadores e artistas que podem não ter o devido reconhecimento no processo de produção de arte generativa. Denomino essa propriedade de “visceralidade”. O artigo conclui que a arte generativa não sucede em criar uma arte não antropocêntrica por causa do fator visceralmente humano presente na inteligência artificial.

PALAVRAS-CHAVE: arte generativa; arte e inteligência artificial; economia da cultura; economia da inteligência artificial; doença dos custos

Artificial surface, visceral interior: Technological, aesthetic and economic aspects of generative art

ABSTRACT The dissemination of art generated by artificial intelligence, or generative art, is changing the creative industries. This article discusses technological, aesthetic and economic issues of generative art. Although technology can make possible new forms of artistic expression, its status as art is not certain. Defenders see it as a non-anthropocentric artform, critics emphasize that art is made by humans. Cultural economics can understand economic aspects of generative art, such as the decision of adopting it in large scale to stave off the cost disease; it also shows that there is a chain of workers and artists that may not be given the due recognition in the production process of generative art. I call this trait of “viscerality”. The article concludes that generative art does not succeed in creating non-anthropocentric art because of the viscerally human factor in artificial intelligence.

KEYWORDS: generative art; arts and artificial intelligence; cultural economics; economics of artificial intelligence; cost disease
Introdução

* Universidade Federal de Minas Gerais

Introdução

¹ Um cover da música “Boy’s a liar” da rapper inglesa PinkPantheress cantado pelas vozes dos presidentes americanos Joe Biden e Barack Obama tem mais de três milhões de visualizações, o que é um número consideravelmente alto para o YouTube (<https://www.youtube.com/watch?v=ou-qk4MwqO8g>). Este é apenas o mais visualizado dos inúmeros exemplos na plataforma.

Imagine um futuro onde imagens geradas por programas de computador estão fixadas nas paredes de museus ou adornando casas de colecionadores de arte e pavilhões artísticos em prédios corporativos; livrarias estão cheias de best-sellers como *A Odisseia* no estilo de Machado de Assis, o livro final das *Crônicas de Gelo e Fogo*, completado automaticamente a partir de notas do autor inicial e um livro de autoajuda escrito por um robô sobre a experiência de ser um robô; no topo dos serviços de streaming, há novas músicas de Cazuza e Rita Lee, junto com um cover de Lampião cantando “Faroeste Cabloco”; um filme de animação escrito por um programa, animado por outro, dublado por vozes geradas por um programa concorre ao Oscar de melhor animação. Neste futuro, todas essas criações precisaram apenas de um número mínimo de trabalhadores para encontrarem o melhor código possível nas plataformas de apoio aos desenvolvedores, gerarem os prompts e retocarem quaisquer problemas.

Embora este cenário não passe de uma fantasia (na data de escrita), todas essas possibilidades já existem em certo grau: uma galeria de arte gerada por inteligência artificial, ou arte generativa, abriu em março de 2023 nos Países Baixos (Katanich, 2023); livros escritos artificialmente inundam a Amazon (Roscoe, 2023); existem tutoriais de como criar qualquer cover com qualquer voz¹ (How to make..., 2023); numa parceria entre o estúdio Fox e a IBM foram usados 100 trailers de filmes de suspense e terror para criar cenas geradas artificialmente para um filme fictício chamado Morgan (Inteligência artificial..., 2016).

Por meio de um cenário imaginado contrastado com a realidade, podemos ver que a inteligência artificial (IA) está penetrando no campo artístico. A explosão de dados disponíveis com a popularização da internet foi fundamental para o desenvolvimento de tecnologias de IA, pois elas precisam de bases de dados vastas para o treinamento dos algoritmos (Albuquerque, 2019; KRÄMER, 2022). O potencial da IA em reduzir custos de serviços mais complexos pode auxiliar na lucratividade das empresas e gerar maior satisfação dos consumidores (Pinheiro; Bastos, 2021).

Tanto Krämer (2022) quanto Metcalfe et al (2021) argumentam que é errado perceber a relação entre humanos e IA num dilema binário entre a substituição completa do humano por máquina e subestimar as capacidades da IA de reproduzir ações huma-

nas, selecionando certos campos de ação humana considerados impossíveis de serem substituídos – o campo artístico, como visto, é um desses campos que se pensavam intocáveis. Por isso, um tratamento mais cauteloso de suas questões e desenvolvimentos é necessário. O objetivo desse artigo é verificar o impacto da IA no campo artístico usando a economia da cultura para analisar questões econômicas da IA.

A definição de IA para esse artigo segue Metcalfe et al (2021, p. 70242): “o domínio preocupado com agentes inteligentes que têm capacidade sensoriais, perceptivas, raciocínio rudimentar e de aprendizado”. Essa definição enfatiza a capacidade de crescimento autônomo de programas e códigos, em que a interferência humana é minimizada ou até eliminada – o processo comumente chamado de machine learning. Tal esclarecimento é necessário porque muitos estudos igualam IA com robotização, sendo que ambos os conceitos não são necessariamente os mesmos (Yingying; Yixiao, 2021). Nesse sentido, o campo artístico também está experimentando um choque tectônico devido à introdução de IA.

O primeiro grande choque aconteceu quando o quadro Edmond de Belamy, criado por meio de processos de machine learning, foi vendido em 2018 pelo lance de US\$432.000; a casa de leilões Christie’s esperava arrecadar até US\$10.000. O resultado foi considerado um sinal de que imagens geradas por IA podem ser a próxima evolução na arte (Arbiza, 2020). Isso obviamente gerou polêmicas, especialmente sobre o papel do artista nesse novo arranjo econômico. Jason Allen, ao ganhar o prêmio de uma feira de arte no Colorado usando uma imagem gerada no programa Midjourney, declama: “A arte está morta, cara...A inteligência artificial ganhou. Os humanos perderam”. (Vallance, 2022).

Embora Allen esteja sendo dramático, suas palavras refletem temores populares gerais de que a inteligência artificial pode substituir seres humanos. John Maynard Keynes (1930 p. 364) definiu isso como o “desemprego tecnológico”, que acontece quando “a nossa descoberta de meios de economizar trabalho acontece num ritmo maior do que novos usos para o trabalho.” Estudos econômicos tendem a considerar IA como um substituto para seres humanos, o que pode gerar desigualdade no curto prazo, devido à necessidade de adaptação (Yingying; Yixiao, 2021). Assim, o setor criativo também está sujeito a essas mesmas preocupações.

2 Deve ser notado também que existem várias perspectivas econômicas, mas a dominante – incluindo em economia da cultura – é a neoclássica, a qual esse artigo aborda e procura demonstrar suas limitações. Para uma introdução a diferentes abordagens de economia da cultura, ver Autor (2019).

Os exemplos anteriores também ilustram o fato de que a esfera econômica e a esfera artística estão intimamente relacionadas. Como colocou Terry Eagleton ([1976] 2011, p. 107, ênfase adicionada), crítico literário britânico, “A literatura[, a arte em geral,] pode ser um artefato, um produto da consciência social, uma visão de mundo; mas ela é também uma indústria”. A economia da cultura – ou economia das artes e literatura – trabalha com esses aspectos, com a relação das artes com a economia (aspectos financeiros da arte, oferta e demanda de produtos artísticos, etc.) e da aplicação do método econômico – associado à escola neoclássica² – à arte (o que determina a decisão de se virar um artista, se o método econômico pode contribuir para discussões sobre a natureza da arte, etc.) (Frey, 2019, p. 3). Uma das ideias mais fundamentais da economia da cultura é a “doença do custo” (Baumol; Bowen, 1966; Besharov, 2005; Baumol, 2012) e o artigo utiliza esse conceito para explicar porque, especialmente em ambientes corporativos, pode existir grande demanda por arte generativa. Devido a diferenças de como artistas, engenheiros e economistas entendem arte, o artigo argumenta que a expansão das artes generativas pode exacerbar conflitos de interesse entre diferentes atores da indústria cultural, expondo a visceralidade da IA.

Entrelaçamentos entre arte e tecnologia

Embora possa não parecer óbvio num primeiro momento, a história da arte e a história da tecnologia se cruzam muitas e muitas vezes. Um exemplo icônico é o da Renascença. Loch, (2021, p. 44-45) escreve que “Brunelleschi [arquiteto responsável pela construção da Catedral de Florença] fez... escolhas conscientes das tecnologias de seus tempos para aprimorá-las e conseguir executar sua obra”. A noção de perspectiva, fundamental para artistas hoje, foi desenvolvida durante esse momento, para colocar em prática uma arte matemática (Oliveira; Moreira, 2014). Assim, os artistas renascentistas eram também “inventores, cientistas, engenheiros, arquitetos, médicos”, tudo para retratar o ser humano da melhor forma possível em suas obras (Loch, 2021, p. 47).

Outro exemplo importante da transformação da arte com novas tecnologias vem da fotografia. Por um bom tempo, desde a primeira fotografia, em 1839, ela era vista como uma violação da estética vigente, influenciada pelo renascimento e pela arte barroca, de que a arte não deveria ser uma transposição direta do real para a tela; com a fotografia,

tal distinção não era mais necessária (Ramos, 2009). Houve uma grande resistência ao reconhecimento da arte fotográfica – Walter Benjamin ([1935] 2017, p. 68) chegou a dizer que, no máximo, a foto pode gerar uma “performance artística”. Até hoje procura-se distinguir uma foto “comum” de uma “artística”, mas seu status como arte é aceito.

No mundo contemporâneo, o nível de intensidade tecnológica dos produtos artísticos varia de caso para caso numa magnitude impensável no passado. Museus continuam apresentando obras de artesãos que usam ferramentas simples, juntamente com outras que utilizam impressoras 3D. Aparelhos de gravação e reprodução de imagens, por exemplo, são classificados como produtos de alta intensidade tecnológica (Teixeira; Missio, 2021, p. 155); isso significa que um filme é um produto de ponta, não somente por seu valor tecnológico, mas pela quantidade de serviços que até mesmo um filme independente de baixo orçamento precisa mobilizar para chegar ao consumidor final.

Nesse contexto, a IA emerge como mais uma forma de transformar as artes. Além dos exemplos da introdução, num relatório do Fórum Econômico Mundial (2018), os autores preveem que a IA poderá gerar uma música original no top 40 em 2027, vídeos criativos em 2028 e um best-seller em 2049. Colocam-se, então, expectativas econômicas enormes, que Caramiaux e Alaoui (2022, p. 2) classificam de “fantasia”. Mas, assim como as tecnologias que a precederam, a IA pode ser um “amplificador da imaginação”, nas palavras de Harold Cohen, que permitirá aos artistas verem “além” do que podem (Hertzmann, 2018, p. 15)

Historicamente, Cohen é considerado o primeiro artista a usar inteligência artificial, com seu programa Aaron, de 1973. Cohen já era um pintor conceituado e projetou Aaron para ser um pintor automático. Desde então, inúmeros artistas têm explorado a interface entre IA e novas tecnologias (Audry, 2021; Audry; Ippolito, 2019; Camariaux; Alaoui, 2022; Grba, 2022). Audry (2021) reúne análises de dezenas de obras de arte feitas com o auxílio de tecnologias de IA, enquanto dissectiona o processo técnico por trás delas. No fim, “espécies diferentes de machine learning permitem tipos diferentes de práticas artísticas e qualidades estéticas” (ibid., p. 18). Artistas generativos se veem como herdeiros da arte conceitual, inaugurada por figuras como Marcel Duchamp (Audry; Ippolito, 2019; Carrasco, 2022; Grba, 2022)

Anatrasirichai e Bull (2021) fazem uma revisão de literatura sobre os avanços da inteligência artificial nas indústrias criativas, dividindo suas aplicações em cinco categorias principais: (i) criação de conteúdo, (ii) análise de informação, (iii) melhoria de

3 Antrasirichai e Bull (2021, p. 591, 636) citam uma pesquisa da Adobe feita com artistas dos Estados Unidos, Reino Unido, Alemanha e Japão sobre a aceitação de ferramentas de IA em áreas como “procura de imagens, edição e outras tarefas ‘não criativas’” que exigem esforço repetitivo como “fazer buscas em uma grande base de dados e a examinar para tirar conclusões”. Mais de três quartos dos correspondentes aceitavam o uso de IA nesses casos.

4 A ideia da IA como algo “alienígena” é defendida por vários pesquisadores de IA, como Nick Bostrom e Eliezer Yudkowsky (LANDGREBE; SMITH, 2022, p. 48).

conteúdo e pós produção, (iv) melhoria e extração de informação e (v) compressão de dados. Embora haja certo consenso³ entre os artistas quanto à utilidade das aplicações (ii), (iv) e (v) em processos auxiliares ao criativo, as categorias (i) e, até certo ponto, (iv) são muito mais controversas. Elas não são simplesmente auxílios, mas têm por intenção criar um possível produto final, capaz de ser considerado “arte” (Mazzone; Elgammal, 2019; Arbiza, 2020; Barale, 2021; Carrasco, 2022).

Edmond de Belamy pertence à categoria (i) – julgando pelo valor pecuniário apenas, a gravura demonstrou que é possível uma IA competir com um artista humano, um campo no qual se assumia que inteligência humana não poderia ser substituída por IA (Metcalf et al 2021, p. 70245). Cetinic e She (2022, p. 15) escrevem que vários defensores da arte generativa argumentam que trabalho produzido por sistemas de IA é “indistinguível de arte feita por seres humanos e é percebida como surpreendente, interessante ou esteticamente prazerosa por um número grande de pessoas”. “Não conseguimos parar de ver elas”, escreve Barale (2021, p. 203) em seu panegírico à arte generativa. Leonel Moura, artista que usa robôs móveis que fazem desenhos gerados aleatoriamente, baseados nas densidades das cores, acredita que não é importante se a arte é feita com ou sem máquinas, desde que “seja aceita pelos guardiões do mundo artístico” (Audry; Ippolito, 2019, p. 2).

Para os defensores, ela se torna, então, uma nova forma de fazer arte, “uma síntese de conhecimento do mundo e invenção”, em que sua engenhosidade depende de “identificar novas formas de combinar e extrapolar os dados” (Kalpokas, 2023, p. 3). Novos desenvolvimentos na pesquisa de IA promovem a ideia de não somente considerar a IA como uma ferramenta, mas uma “parceira”, onde tanto o usuário humano quanto a IA em si mesma melhoram um ao outro (Metcalf et al, 2021). Por isso, vários autores consideram que a arte generativa é válida como uma arte única, não antropocêntrica (Carrasco, 2022; Kalpokas, 2023)⁴. Schober (2022, p. 159) chega a classificar de “capacitista” a ideia de que a criatividade só está ligada a habilidades cognitivas e à inteligência humana. Assim, cria-se expectativas sobre a capacidade das IAs de imitar seres humanos para que um dia elas possam ser singularmente separadas destes.

O significado da arte na era da supremacia tecnológica

Nos exemplos citados acima – arte e técnica renascentista, fotografia e, eventualmente, arte generativa – deve se ter em mente que discussões em torno de tais processos evoluíram ao decorrer de anos, décadas. É um processo fluído que não está e, provavelmente, nunca vai estar finalizado. Por isso, a questão “o que é arte?” não terá uma resposta definitiva e não ambígua. “Arte com A maiúsculo,” escreve Gombrich ([1950] 1989, p. 15) na introdução de sua história da arte, “não existe...[ela] é um fantasma e um ídolo”. Isso cria um problema, porque artistas tendem a ter diferentes objetivos em relação aos engenheiros que programam os códigos e estruturas e aos economistas que estudam as relações de mercado e produção.

Quando reduzida a mera técnica, arte é repetição. O que é uma pintura senão pinceladas contínuas e calculadas numa tela? O que é uma novela senão uma concatenação de tropos em uma ordem compreensível ao leitor? O que é música senão um quebra-cabeça de fonemas e frequências, que pode ser representado numa partitura? Quando vistos dessa forma, produtos artísticos podem ser entendidos como fruto de um processo de repetição e otimização. Eles são reprodutíveis (Benjamin, [1935] 2017) e, portanto, potencialmente programáveis para a reprodução numa máquina. David Chalmers (2010, p. 14), filósofo de IA, argumenta que o cérebro é uma máquina, “um sistema complexo composto de partes governadas por leis interagindo de uma forma governada por uma lei” e, por ser uma máquina, existe a possibilidade de que uma inteligência artificial geral (IAG) – capaz de dar igualdade à máquina em relação ao ser humano – deixe de ser ficção científica. Dessa forma, podemos derivar “leis” da arte que podem ser algoritmizáveis, para que uma IA possa criar arte independente de humanos, eventualmente. Nas palavras de Arbiza (2020, p. 58), a criatividade “pode ser investigada, simulada e extraída para o bem da sociedade”, esperando por um futuro onde uma IAG possa “dar significado ao seu próprio trabalho”.

O projeto The Next Rembrandt, de 2016, patrocinado pelo banco holandês ING e a Microsoft usou um sistema de deep learning que analisou 346 obras de Rembrandt para criar uma pintura “inédita” do artista holandês. O projeto levou mais de 18 meses para ser completo, envolvendo mais de 20 profissionais de computação, análise de da-

dos, design, entre outros. O resultado foi uma obra supostamente indistinguível para o leigo (Loch, 2021, p. 73-75). O experimento parece confirmar a ideia de que arte é um processo de repetição, reproduzível e, portanto, algoritmizável. Como Ballietti (2020) argumenta, se a beleza obedece a padrões matemáticos, logo podemos sintetizar as principais características das obras nos grandes arquivos em “leis da arte”, extrair as proporções matemáticas mais belas e agradáveis e ensinar IAs a reproduzi-las. Castellano, Lella e Vessio (2021) veem um grande potencial da IA para a história da arte, ao verificar padrões de semelhança nas pinturas de grandes artistas e, assim, analisar a influência que um exerceu sobre o outro; os resultados podem ser potencialmente alimentados a IAs. Assim, o trabalho humano é economizado e custos são diminuídos: não teremos somente novos Rembrandts, mas novos Da Vincis, novos Picassos, etc.

Alguns programas como Stable Diffusion e Midjourney podem gerar desenhos usando um prompt de texto (por exemplo, “Sapo num monociclo ao estilo de Picasso”). Sanchez (2023, p. 57) observa que “nomes de artistas” são o componente mais comum utilizado. Embora Sanchez observe que a maioria dos usuários são praticantes de hobby, tal praticidade eventualmente atrairá investidores. Como explorado em Antrasirichai e Bull (2021), um dos principais usos da IA é procurar e catalogar informações, incluindo as de mercado (preferências do consumidor, preços de materiais) – potencialmente, uma IA avançada o suficiente pode combinar várias funções e criar um produto baseado nelas.

Porém, tal entendimento da arte entra em conflito com as visões dos artistas. Sofian Audry, um artista que tem formação em matemática, computação e humanidades, relata que seu primeiro contato com o mundo da arte foi com uma visão borrada por preconceitos, em 2007:

Meu projeto [de arte computacional] iria criar uma solução ótima ao “problema” da criação artística e poderia potencialmente substituir os artistas de uma vez por todas. Essa ideia incorporou um certo grau de desprezo por meus colegas artistas, que se formaram em escolas de arte ou de cinema e (porque eles não possuíam o conhecimento e as habilidades que eu tinha) certamente não poderiam competir com minha abordagem brilhante... Nada poderia estar mais além da verdade. Fazer arte é desafiador. Fazer arte boa requer não somente intuição e talento, mas também dedicação inabalável e uma forte dose de resiliência. (Audry, 2021, p. 22)

O relato de Audry demonstra um problema comum da relação da arte com o público: a ignorância em relação ao esforço na produção artística. Como um forasteiro, Audry tinha também esse entendimento falho do que era a arte, o qual ele se desencilhou à medida que foi aprendendo sobre o mundo artístico.

Existem visões diferentes sobre o “problema da arte”. Como enfatiza Hannah Arendt, a arte não tem ordem útil, ou seja, não se cria arte com um objetivo de uma aplicação imediata (apud Loch, 2021, p. 34, 70). Não é um “fato mensurável”, mas uma “condição interpretada” (Audry; Ippolito, 2019, p. 7). O problema da engenharia de IA é de otimização, como tornar processos mais eficientes (Audry, 2021). Tal problema é semelhante ao de economia, em que, especialmente após a Segunda Guerra Mundial, economistas viam o problema do crescimento econômico como um problema de uso ótimo dos recursos naturais e humanos (Backhouse, 2009; Boianovsky; Hoover, 2014). Besharov (2005) lamenta que a teoria econômica não conseguiu encontrar uma explicação econômica para a excepcionalidade da arte, mas, enquanto a economia (neoclássica) adotar o paradigma otimizador, a arte vai ser apenas um produto⁵. Ambos os processos otimizadores de engenharia e economia não conseguem capturar a “aura” (Benjamin, [1935] 2017) presente nas obras de arte.

Se a arte não tem ordem útil, significa que pode não fazer sentido “otimizá-la”, tanto do ponto de vista engenheiro quanto o econômico. Processos de otimização podem fazer parte da produção artística – como num levantamento de custos para a obtenção do melhor preço possível para a matéria-prima de uma escultura –, mas não podem determinar se o produto final é arte ou não.

O relato de Audry acima é relevante também porque as ideias do Audry de 2007 sobre arte são mais comuns do que se supõe. Retornando a Edmond de Belamy, embora Arbiza (2020, p. 49) coloque o coletivo francês Obvious, que estava por trás dos códigos usados para o produto, como defensor da democratização da arte, críticos profissionais declararam a obra pouco original e desinteressante, um exemplo pobre da arte generativa (Cetinic; She, 2022; McCormack, 2022). McCormack (2022, p. 184) também observa que os membros do coletivo não tinham nenhum treinamento ou reputação em arte. Muitos dos defensores da arte generativa que se julgam artistas conceituais, escreve Grba (2022, p. 18), ignoram o quanto a arte evoluiu além de Duchamp. Comentando sobre The Next Rembrandt, Loch (2021, p. 78) escreve que é mais

⁵ Baumol e Bowen (1966, p. 162) reconhecem que a arte é também “uma manifestação intangível do espírito humano”, mas para se fazer uma análise econômica, ela deve ser tratada como “uma atividade produtiva que fornece serviços à comunidade”. Este tem sido o paradigma de estudos empíricos relacionados à arte.

um produto “sobre o fetiche da máquina substituir cada vez mais seu criador, que aqui parece ser a intenção por parte da Microsoft: mostrar os potenciais de uma máquina e não propor uma poética.”

Mesmo que a IA gere produtos esteticamente agradáveis, isso não implica que eles sejam arte; paisagens naturais evocam sentimentos de beleza e do sublime, criando conexões com a “audiência”, mas não foram criadas por seres humanos (Hertzmann, 2018). Sobre isso, Natale e Henrickson (2022, p. 2) chamam de “efeito Lovelace”, que ocorre quando “o comportamento de sistemas computacionais é percebido pelos usuários como original e criativo”. Quando Barale (2021, p. 209) escreve que o que nos atrai para a arte generativa é sua “luta por identidade e como elas mostram seu esforço; as incertezas e as possibilidades abertas que caracterizam o momento”, ela demonstra estar antropomorfizando a arte generativa, num exemplo claro do efeito Lovelace. Assim, a crítica de Benjamin ([1935] 2017, p. 68) à fotografia pode ser adaptada a vários exemplos de arte generativa: é uma performance; o que gerou a arte foi “o arranjo institucional que estava ao redor da sua apresentação ao público” (Natale; Henrickson, 2022, p. 11). Mazzone e Elgammal (2019), em sua defesa da arte generativa, admitem que o ambiente da exposição e o jogo com o público se a figura fixada na parede é artificial ou não, também fazem parte da performance artística; embora eles defendam que a figura em si é arte, ela precisa do contexto de sua performance e tal contexto pode ser construído para se aproveitar dos vieses de decisão para induzir ao resultado esperado (Kahneman, 2011). Em experimentos fora das galerias de arte generativa, observou-se rejeição, considerando-a inferior (Ragot; Martin; Cojea, 2020; Bellaiche et al 2023; Moura; Castrucci; Hindley, 2023). Grba (2022, p. 11, 12) observa que muitas das instalações de arte generativa são “hiperesteticizadas” e que muitos trabalhos ruins podem ser compensados por “altos valores de produção” por meio do poder corporativo do dinheiro, “obscurecendo práticas experimentais e avant-garde”.

Por isso, em sua crítica à arte generativa, Hertzmann (2018, p. 2) enfatiza que “computadores não criam arte, pessoas usando computadores criam arte”. Ele concorda com a ideia de que a IA fornece ferramentas úteis, conforme as categorias (ii), (iii), (iv) e (v) de Anatrasiachai e Bull (2021), mas defende que a categoria (i) não é arte por si só. Ao adotar a ideia de que a arte é uma “interação entre agentes sociais”, ele está incluindo os agentes que não somente podem produzir obras que podem ser consi-

deradas arte, mas que têm capacidade de usar a arte para expressar empatia e considerações éticas. Mesmo que um programa possa produzir uma pintura esteticamente prazerosa, não será muito diferente de uma paisagem natural, nesse sentido.

Ao adotar essa perspectiva de arte, é possível também criticar a ideia de que artistas são “máquinas de intenção” (Hertzmann, 2018, p. 18). De fato, Hertzmann viu perigos em chamar programas de “artistas”, porque se está atribuindo emoções e capacidade de julgamento humanas – o efeito Lovelace – a entidades que definitivamente não as têm. Isso pode “sobrestimar a competência e habilidades da IA” e “levar ao engano sobre a natureza da arte” (ibid., p. 21). McCormack (2022, p. 188) concorda com isso, pois o “software de IA não tem intenção ou autenticidade” e que chamar o que eles produzem de “arte” seria esvaziar o significado da arte.

A ideia de que artistas não são “máquinas de intenção” também serve como crítica à ideia de que o cérebro é uma máquina. Landgrebe e Smith (2022) argumentam que a ideia de se tratar o cérebro (e o corpo humano, em geral) como máquina é errônea porque é um sistema complexo e, portanto, algo que não pode ser modelado até seus pormenores. Tal entendimento vem do dualismo cartesiano, da separação entre corpo e mente, algo que está sendo cada vez mais abandonado – em outras palavras, considera-se que há um contínuo entre o corpo e a mente, e isso não pode ser modelado e representado em lógica simbólica, conforme nos parâmetros de Turing. Isso torna a algoritmização impossível, assim impossibilitando uma IAG. Falta um modelo de impulsividade⁶, que permita a IAs fazerem escolhas realmente independentes, que podem escapar dos seus ambientes controlados, para que elas possam realmente crescer. Logo, Landgrebe e Smith (2022, p. 207) argumentam: “A IA é, no fim das contas, apenas um conjunto de modelos de Turing matematicamente computáveis que tomam inputs e computam outputs de forma determinística. E é assim mesmo quando há humanos que acreditam que um computador é consciente”.

Seguindo essa linha de argumentação, é também impossível para uma IA produzir arte, porque não há um modelo formal que responda à pergunta “o que é arte?” Ou seja, não há uma equação que decida o que é arte e nem é o desejo da vasta maioria dos artistas que ela exista. Isso é perseguir um fantasma, nos termos de Gombrich. Falta, então, uma interação social real entre o computador e o resto (Hertzmann, 2018). Assim, o que a IA pode fazer é produzir um produto que pode ou não ser considerado

6 No original, *drive-ness*. Landgrebe e Smith (2022, p. 300) definem *driveness* como “a propriedade de um sistema de não convergir ou se mover em direção a um equilíbrio [...], mas de um estado para outro como resultado de uma força exercida interna ou externamente, que pode ser contínua, regular ou pontual. Por exemplo, a contração do coração é um exemplo da aplicação de uma força pontual interna regular dentro do sistema cardiovascular.”

arte ou performance artística por seres humanos. Mas ela não vai produzir arte por si mesma. A arte continuará sendo um processo visceralmente humano.

A visceralidade da máquina: relações econômicas por trás da arte generativa

Análises mais equilibradas do impacto da IA no mundo artístico entendem que pode existir uma parceria entre o ser humano e a máquina; existe espaço para a arte generativa, apesar dos problemas indicados. Por outro lado, a ideia de que a arte generativa criará uma arte não antropocêntrica está mais próxima da visão fantasiosa da IA do que da realidade. Humanos ainda coletam e selecionam os dados a serem computados, eles que programam os algoritmos, eles que organizam as exposições e a divulgação, eles que definem “arte”. Além disso, por trás dos processos generativos, há sempre pessoas como “mechanical turks” (Landgrebe; Smith, 2022, p. 292). “inteligência artificial artificial” ou “pseudo-IA”, que normalmente são mal pagos (GRBA, 2022, p. 7). Novamente, McCormack (2022, p. 184) observa que os membros do coletivo Obvious, para produzir Edmond de Belamy, apropriaram-se de um código open source, cujo autor – Robbie Barat (que por sua vez se baseou em outros códigos) – não foi creditado, nem remunerado. Edmond de Belamy foi apresentado como produto de uma IA e isso pode induzir sua audiência ao engano de que não houve interferência humana na sua criação (Hertzmänn, 2018; Cetinic; She, 2022).

A questão da remuneração dos artistas é algo que preocupa os economistas. De fato, a economia da cultura teve sua fundação na análise dos custos nas artes performáticas (orquestra, teatro, ópera, dança) (Baumol; Bowen, 1966; Besharov, 2005; Baumol, 2012). Nesse sentido, adota-se a ideia de que uma das principais fontes de crescimento econômico vem dos ganhos de produtividade causados pela substituição de trabalhadores por máquinas. Porém, como Baumol e Bowen (1966) perceberam, isso não ocorre no setor de artes performáticas, porque a natureza desses bens faz com que não tenham os mesmos ganhos de produtividade da economia. A 9^ª Sinfonia de Beethoven dura 80 minutos e precisa de 19 músicos; não faz sentido investir numa tecnologia para fazer com que ela dure 40 minutos e necessite de 9,5 músicos, mes-

mo que custe metade do original. Em outras palavras, mesmo que haja melhorias nas tecnologias adjacentes, como efeitos especiais e transporte⁷, aumentos de produtividade não ocorrem diretamente nos produtos artísticos. Posteriormente, essa análise foi estendida para outros tipos de arte e, mais amplamente, para os “serviços pessoais”, como saúde e educação; Baumol (2012) usa esse conceito para explicar o crescimento explosivo dos custos destes nos Estados Unidos, em contraste com a queda nos custos de produtos como computadores e televisores. Dá-se o nome a esse fenômeno de “doença dos custos”. E ela é uma doença crônica.

Logo, a produtividade no setor cultural não vai crescer na mesma proporção que o resto da economia, fazendo com que os salários dos trabalhadores envolvidos – especialmente artistas – não acompanhe o do resto. Do ponto de vista econômico, isso cria um desincentivo para as pessoas ingressarem na indústria cultural, que pode ser acompanhado de estigmatização do trabalho do artista. Ocorre uma maior probabilidade de precarização dos trabalhadores da indústria cultural, que dá origem ao estereótipo do “artista pobre” (Baumol; Bowen, 1966).

Devido ao excesso de oferta e racionamento de empregos, o mercado para artistas tende a ser altamente competitivo (Menger, 1999; 2006). Além disso, o setor cultural tende a ser dominado por “superestrelas”, que concentram o mercado, tendo recursos o suficiente para defender seus direitos autorais e viver de sua arte (Adler, 2006). Isso cria um ambiente altamente desigual, tanto do ponto de vista de remuneração, representação e definições artísticas (Baumol, 2006). Devido a essas características, é necessário o artista ter paixão pelo trabalho, algo que tende a ser explorado por produtores menos escrupulosos, onde salários menores são oferecidos para pessoas que aceitariam trabalhar por menos porque estão fazendo o que amam (Tojumitsu, 2015). A maioria dos artistas tendem a ter mais de um emprego, para que possam sustentar sua arte não só como hobby, mas também como profissão (Throsby; Zednik, 2011). Frey (2019, p. 37) observa que artistas tendem a ser mais satisfeitos do que não artistas, devido à liberdade do seu trabalho, mas tal satisfação está ligada à estabilidade financeira; usando exemplos empíricos, tanto atores de teatro alemão (Eikhof; Hauschild, 2007) quanto músicos de Belo Horizonte (Henrique; Machado; Antigo, 2022) têm um alto nível de paixão pelo trabalho, mas que só é eficazmente desfrutado quanto maior for a estabilidade financeira (cf. Baumol; Bowen, 1966).

7 Nesse sentido, a queda nos custos do papel permitiu que o número de pintores, escritores, músicos crescesse vertiginosamente porque diminuí custos de prática (qualquer artista tem cadernos e cestas de lixo cheias de esboços), pois não há mais a preocupação em economizar papel (ver BAUMOL, 2012, p. 37).

Mesmo que novas tecnologias sejam apresentadas como uma solução para a doença dos custos, pode não ser o suficiente. O exemplo mais relevante vem da indústria cinematográfica. A construção de tal indústria foi possível graças ao desenvolvimento tecnológico, que permitiu o surgimento de uma nova forma de arte, e a melhoria dos canais logísticos e de acesso do público a produtos artísticos (Baumol, 2006). Porém, a indústria de artes gráficas – que realizam os efeitos especiais de filmes e séries, sendo um dos principais sustentáculos de Hollywood – tem seus trabalhadores sujeitos a longas horas de trabalho, remuneração baixa e alta instabilidade empregatícia (Salandino, 2022; Iatse, 2023). A indústria cinematográfica como um todo sofre da doença dos custos, refletidos nos custos de produção e marketing, além de incerteza sobre a reação das audiências (Escoffier; McKelvey, 2015), o que leva produtores a tomar cada vez menos riscos e fazer filmes mais previsíveis e de baixo valor artístico (McMahon, 2019). Por isso, o cineasta Martin Scorsese escreve: “Não podemos depender da indústria cinematográfica, do modo como está, para preservar o cinema [como arte]” (apud McMahon, 2022, p. 4).

Logo, o “fetiche da máquina substituir cada vez mais seu criador” tem um objetivo econômico simples: “comercializar suas habilidades em manusear certas tecnologias” (Loch, 2021, p. 78, 99). De um ponto de vista do consumidor, essas tecnologias servem para “promover noções vagas de um estilo de vida consumista aumentado por robôs” (Grba, 2022, p. 7). Do ponto de vista do produtor, tecnologias estão em alta demanda para se escapar da doença do custo. A IA é, certamente, uma dessas tecnologias, pois permite a reprodução em larga escala de processos repetitivos na produção de bens criativos. Na indústria cinematográfica, produtores então se deparam com a seguinte questão: por que eu deveria pagar um escritor humano – que é um empregado antes de ser um artista – se eu posso simplesmente programar uma IA para que faça o mesmo? Deve-se considerar que os grandes estúdios têm acesso a recursos que tornam a utilização de IA potencialmente lucrativa, devido à sua escala (cobrindo custos de energia e hardware para rodar os programas). Por isso, o sindicato Writers Guild of America entrou em greve em 2023, argumentando que o uso arbitrário de IA é prejudicial aos trabalhadores e subverte o próprio significado de arte, o que paralisou vários projetos (Shah, 2023). Novamente há o abismo entre uma visão artística da arte e uma visão econômico-otimizadora da arte.

Como mencionado anteriormente, fora de ambientes favoráveis à IA, as audiências tendem a rejeitar a arte generativa. A razão pela qual obras generativas tendem a ser rejeitadas também está relacionada com o conceito de “vale da estranheza” (uncanny valley). Inicialmente descrito pelo roboticista Masahiro Mori ([1970] 2012), é o efeito do ser humano perceber um robô como se fosse vagamente semelhante a ele, mas também completamente diferente, sinistro e, potencialmente, perigoso. As preocupações de Mori influenciam a criação de robôs e próteses mais agradáveis à visão humana, focando-se menos em realismo e mais em designs estilizados que demonstrem que o robô ou a prótese não estão “tentando” serem humanos, mas entidades independentes. Embora seja a preocupação de vários artistas generativos de enfatizar que a arte gerada por IA não é uma arte “humana”, consumidores podem rejeitar a arte generativa com base no vale da estranheza (Tubadji; Huang; Webber, 2021).

Se a arte é um fenômeno social, ela é mais do que só o produto final. Se a audiência tem uma noção intuitiva disso, ela pode estranhar o fato de não ter sido feita por seres humanos. Mesmo que a audiência reconheça como bonita, ou válida como arte, pode existir algo num produto – especialmente em pinturas generativas – que a deixe sentindo que há “algo de errado”. No início são dedos e mãos a mais, mas, mesmo com a melhoria dos algoritmos para fazer uma reprodução mais fiel do corpo humano, ainda pode existir “algo de errado”. Isso pode ser usado em favor da arte generativa, quando ela explora o “espaço latente”, o “espaço abstrato multidimensional de representações de imagens codificadas”, algo naturalmente estranho e inumano (CETINIC; SHE, 2022, p. 12). Para outros contextos artísticos, o vale da estranheza estética pode fazer com que haja uma rejeição. A baixa qualidade dos efeitos visuais causada pelas condições precárias de trabalho mencionadas acima pode piorar a situação da indústria cinematográfica, além de preocupações com a integridade artística (como “ressurreição” não consensual de atores mortos⁸), perturbação da experiência social do cinema e preocupações com os limites da tecnologia (MURPHY et al, 2023)⁹. Com a tendência dos estúdios de minimizarem riscos, a demanda deles por deepfakes de atores idosos e falecidos pode ter resultados negativos.

A dicotomia arte-como-produto/arte-como-arte se mostra mais relevante na questão da autoria. Como vimos, Edmond de Belamy usou um código, cujo autor não foi creditado. Redes neurais precisam de bases de dados enormes para gerar resultados

8 O filme *The Flash*, de 2023, apresenta várias aparições de atores que interpretaram o Super-Homem em encarnações anteriores, geradas por efeitos gráficos. Um dos mais controversos foi da versão interpretada por George Reeves, na década de 1950. Apesar de ter sucesso como o personagem Super-Homem, foi impossível ele se desvencilhar da imagem do personagem, o que contribuiu para seu suicídio em 1959. Muitos consideraram um desrespeito a Reeves (AYALA, 2023).

9 O filme *O Congresso Futurista* (2012) imaginou um futuro no qual artistas vendem suas imagens para serem utilizadas indefinidamente por estúdios em novos filmes. O filme retrata isso num ambiente no qual a maior parte da população desistiu do mundo real em troca de um mundo virtual, que começou com a digitalização de atores. Na greve de 2023, atores se uniram à greve porque foram oferecidos pagamentos baixos para venderem suas imagens como “réplicas digitais” aos estúdios de forma semelhante ao filme (Buckley, 2023).

10. Aqui entra o problema de tracing, que causa muita controvérsia na comunidade artística, pois trilha uma linha tênue entre referência e plágio (Desmistificando..., 2016). O uso da IA fará com que essa linha fique cada vez mais problemática.

aceitáveis e muitas delas se apropriam de obras que não estão em domínio público. Num exemplo mais extremo, um usuário de IA tirou uma cópia de uma pintura enquanto seu autor fazia um stream (vários artistas usam canais para transmitirem seu trabalho, de forma a se conectarem com sua audiência) e ainda se arrogou o título de autor real da obra (Ge, 2022). Mas várias plataformas de artistas estão usando arte produzida por seus usuários para treinar suas IAs sem consentimento ou com consentimento duvidoso (Junqueira, 2022). Usuários do site Artstation organizaram um protesto fazendo uploads de imagens contra arte generativa, custando ao site acessos (Yalalov, 2023). Como Sanchez (2023) observou, nomes de artistas são os que mais aparecem nos prompts. Porém, esses nomes não são normalmente de artistas clássicos, mas tendem a ser de nichos culturais, como o artista Greg Rutowski, bastante requisitado por sua habilidade em desenhos do gênero de fantasia, cujo nome foi usado 400 mil vezes para gerar imagens no seu estilo sem seu consentimento (Hutchinson; John, 2023). Num mercado altamente competitivo, onde a necessidade de se diferenciar não é somente uma necessidade financeira, mas também está ligada à identidade do artista¹⁰, o potencial reproduzido da IA faz com que a técnica seja dispensável (Mcconnack, 2022). Dessa forma, não são apenas os incentivos econômicos que são danificados (o artista não pode mais ser recompensado por sua arte porque a IA simplesmente a reproduz em escala maior a um custo menor), mas também seus incentivos pessoais (não há incentivo para o artista buscar e explorar inovações, pois a IA pode banalizar o processo de descoberta). Memo Akten, artista generativo, pergunta: “se a máquina cria tudo, então onde está a assinatura dos artistas?” – repare que ele usou o plural, incluindo “pesquisadores que desenvolveram o algoritmo, os programadores que o implementaram, as pessoas que circularam online, as pessoas que geraram os dados para o treinar, os artistas que o apropriaram” (Caramiaux; Alaoui, 2022, p. 16). Podemos adicionar mais uma pergunta: quem está lucrando com isso?

Embora possa se argumentar que é da responsabilidade do usuário estabelecer limites éticos quanto ao uso da IA, Grba (2023) mostra que o mercado de arte digital, especialmente envolvendo non-fungible tokens (NFTs), promove incentivos a comportamentos oportunistas, perversos e até mesmo preconceituosos. Tais preocupações são comumente ignoradas por boa parte da literatura, que se foca na inovação e na possibilidade de arte não antropocêntrica, onde as preocupações dos artistas quanto à

autoria são indeferidas e relativizadas – especialmente quando a maioria dos artistas não são superestrelas e não têm acesso a amplos times legais para se defenderem. Sem considerar esses problemas, as artes generativas falharão como instrumento de democratização, podendo ter o efeito oposto, junto com a multiplicação de espetáculos tecnológicos sem vida.

II Deve se observar que os apoiadores acríticos da arte generativa não são artistas, mas filósofos e pesquisadores com formação em ciências da computação ou engenharia.

Considerações finais

Fazer arte boa é difícil. Qualquer pessoa que diga o contrário não está dizendo a verdade. Assim, fazer arte generativa boa também é difícil. Audry (2021) e Caramiaux e Alaoui (2022) enfatizam que os artistas generativos tentam fazer algo com IA que não é possível ou de forma diferente dos métodos tradicionais. Eles entendem que arte não é neutra, nem mesmo a arte generativa. Eles também não aceitam a ideia defendida por certos interesses de que a tecnologia é neutra, leva automaticamente ao bem maior e que deve ser adotada o mais rápido possível (GRBA, 2022). Longe disso, eles se distanciam da cultura corporativa e otimizadora, buscando escapar de valores como “precisão, produtividade e performance” para explorar as possibilidades da técnica generativa (Caramiaux; Alaoui, 2022, p. 13)¹¹.

Porém, esses valores ainda influenciam muitas decisões corporativas que entram em choque com aquilo que os artistas almejam. Na sua luta eterna contra a doença dos custos, produtores das indústrias culturais certamente verão a IA como uma ferramenta a ser usada, por causa do seu potencial de redução dos custos. Devido aos diferentes objetivos – criação de arte vs. maximização de lucros – haverá conflitos.

Embora a economia não possa definir o que é “arte”, ela pode analisar os processos humanos por trás dela, incluindo sua visceralidade. Assim como não é correto antropomorfizar a máquina, não é correto maquinizar o ser humano, e isso também vale para o campo artístico. A ideia de uma arte não antropocêntrica ainda está longe de ser realizada. O discurso que promove o não antropocentrismo da arte generativa tende a minimizar as relações humanas por trás dela. Suas promessas de democratização serão vazias se não considerarem isso.

Porém, a arte generativa faz novas questões não só sobre a natureza da arte, seus custos e sua autoria, mas também da nossa relação com o consumo – podemos consumir arte *como* arte ou ela é *apenas* um serviço, não diferente de um almoço num restaurante? A natureza da satisfação é a mesma? Retornando às palavras de Allen na introdução, se “os humanos perderam”, eles não perderam para as IAs, mas para interesses *visceralmente* humanos por trás delas. Por isso, entender melhor a arte e como criar parcerias com as ferramentas de IA enquanto se respeitam os limites deve ser o objetivo de novas pesquisas. Apesar do termo “desemprego tecnológico” ser negativo, Keynes (1930) acreditava que a Grande Depressão, vigente quando o ensaio foi publi-

cado, iria passar; novas tecnologias iriam melhorar o mundo e tornar as artes mais acessíveis. Com um uso mais consciente, a IA pode ajudar a cumprir esse objetivo.

Referências

- ADLER, M. *Stardom and talent*. In: GINSBURGH, V. A.; THROSBY, D. (eds.). *Handbook of the economics of arts and culture*. Amsterdam: North Holland, 2006, p. 895-907.
- ALBUQUERQUE, E. M. *Revoluções tecnológicas e general purpose technologies: mudança técnica, dinâmica e transformações do capitalismo*. In: RAPINI, M. S. et al (eds.). *Economia da ciência, tecnologia e inovação: fundamentos teóricos e a economia global*. Belo Horizonte: CEDEPLAR, 2021, p. 53-83.
- ANATRASIRICHAI, N.; BULL, D. *Artificial intelligence in the creative industries: a review*. *Artificial Intelligence Review*, n. 55, p. 589-656, 2022. DOI: 10.1007/s10462-021-10039-7.
- ARBIZA G., M. *A critique of contemporary artificial intelligence art: who is Edmond de Belamy*. *AusArt: Journal for Research in Art*, v. 8, n. 1, p. 49-64, 2020. DOI: 10.1387/ausart.21490.
- AYALA, N. *The Flash's controversial black & white Superman cameo: who was George Reeves?* *Screen-Rant.com*, 18 jun. 2023. Disponível em: <https://screenrant.com/the-flash-movie-george-reeves-superman/> Acesso em: 5 ago. 2023.
- AUDRY, S. *Art in the age of machine learning*. Cambridge: MIT Press, 2021.
- AUDRY, S.; IPPOLITO, J. *Can artificial intelligence make art without artists? Ask the viewer*. *Arts*, v. 8, n. 35, 2019. DOI: 10.3390/arts8010035.
- BACKHOUSE, R. E. *Economics*. In: BACKHOUSE, R. E.; FONTAINE, P. (eds.). *The history of social sciences since 1945*. Cambridge: Cambridge University Press, 2009, p. 38-70.
- BALLIETI, S. *The human quest for discovering mathematical beauty in arts*. *Proceedings of the National Academy of Sciences*, v. 117, n. 44, p. 27073-27075, 2020.
- BARALE, A. *"Who inspires who?" aesthetics in front of AI art*. *Philosophical Inquiries*, v. 9, n. 2, p. 195-220, 2021. DOI: 10.4454/philing.v9i2.367
- BAUMOL, W. J. *The arts in the "new economy"*. In: GINSBURGH, V. A.; THROSBY, D. (eds.). *Handbook of the economics of arts and culture*. Amsterdam: North Holland, 2006, p. 339-358.
- BAUMOL, W. J. *The cost disease: why computers get cheaper and health care doesn't*. New Haven: Yale University Press, 2012.
- BAUMOL, W. J.; BOWEN, W. G. *Performing arts: the economic dilemma*. New York: Twentieth Century Fund, 1966.
- BELLAICHE, L. et al. *Humans versus AI: whether and why we prefer human-created compared to AI-created artwork*. *Cognitive Research: Principles and Implications*, v. 8, n. 42, p. 1-22, 2023. DOI: 10.1186/s41235-023-00499-6.
- BESHAROV, G. *The outbreak of the cost disease: Baumol and Bowen's founding of cultural economics*. *History of Political Economy*, v. 37, n. 3, p. 412-430, 2005. DOI: 10.1215/00182702-37-3-412.
- BOIANOVSKY, M.; HOOVER, K. D. *In the Kingdom of Solovia: the rise of growth economics at MIT, 1956-1970*. *History of Political Economy*, v. 46, supp., p. 198-228, 2014. DOI: 10.1215/00182702-2716172.
- BUCKLEY, S. *Striking actors say rejected 'AI proposal' would let studios use their likeness without fair pay*.

Engadget.net, 13 jul. 2023. Disponível em: <https://www.engadget.com/striking-actors-say-rejected-ai-propo-sal-would-let-studios-use-their-likeness-without-fair-pay-000025878.html>. Acesso em: 5 ago. 2023.

CARAMIAUX, B.; ALAOU, S. F. "Explorers of unknown planets": practices and politics of artificial intelligence in visual arts. *Proceedings of the ACM on Human-Computer Interaction*, v. 6, n. CSCW2, article 477, 2022. DOI: 10.1145/3555578.

CARRASCO B., M. Artistic beauty in the face of artificial intelligence art. In: JOVANOVIĆ, I. V.; STUPNIK, V. M. (eds.). *Social and technological aspects of art challenges of the 'new normal'*. Rijeka: University of Rijeka, 2022, p. 93-112.

CASTELLANO, G.; LELLA, E.; VESSIO, G. Visual link retrieval and knowledge discovery in painting datasets. *Multimedia Tools and Applications*, n. 80, p. 6599-6616, 2021. DOI: 10.1007/s11042-020-09995-z.

CETINIC, E.; SHE, J. Understanding and creating art with AI: review and outlook. *ACM Transactions on Multimedia Computing, Communications, and Applications*, v. 18, n. 2, article 66, 2022. DOI: 10.1145/3475799.

CHALMERS, D. J. The Singularity: a philosophical analysis. *Journal of Consciousness Studies*, v. 17, n. 9/10, p. 7-65, 2010.

DESMISTIFICANDO: a polêmica por trás de NGNL. *Biblioteca Brasileira de Mangás*, 15 set. 2016. Disponível em: <https://blogbbm.com/2016/09/15/desmistificando-a-polemica-por-tras-de-ngnl/>. Acesso em: 5 ago. 2023.

EAGLETON, T. *Marxismo e crítica literária*. São Paulo: Ed. Unesp, [1976] 2011.

EIKHOF, D. R.; HAUNSCHILD, A. For art's sake! Artistic and economic production logics in creative production. *Journal of Organizational Behavior*, v. 28, p. 523-538, 2007. DOI: 10.1002/job.462.

ESCOFFIER, N.; MCKELVEY, B. The wisdom of crowds in the movie industry: towards new solutions to reduce uncertainties. *International Journal of Arts Management*, v. 17, n. 2, p. 52-63, 2015.

FREY, B. S. *Economics of art and culture*. Berlin: Springer, 2019.

GE, J. When AI stole and finished your drawing then calls you a thief. *Superpixel*, 14 out. 2022. Disponível em: <https://www.superpixel.com/article/243730/when-ai-stole-finished-your-drawing-then-calls-you-thief>. Acesso em: 5 ago. 2023.

GOMBRICH, E. H. *La historia del arte*. Ciudad de Mexico: Editorial Diana, [1950] 1989.

GRBA, D. Deep else: a critical framework of AI. *Digital*, n. 2, p. 1-32, 2022. DOI: 10.3390/digital2010001.

GRBA, D. Faux semblants: a critical outlook on the commercialization of digital art. *Digital*, n. 3, p. 67-80, 2023. DOI: 10.3390/digital3010005.

HENRIQUE, J. S.; MACHADO, A. F.; ANTIGO, M.F. Work satisfaction and job permanence in artistic careers: the case of musicians in Belo Horizonte, Brazil. *Journal of Cultural Economics*, 2022. Forthcoming. DOI: 10.1007/s10824-022-09467-2

HERTZMANN, A. Can computers create art? *Arts*, v. 7, n. 18, 2018. DOI: 10.3390/arts7020018.

HOW to make an AI cover song with any artist's voice. *MusicRadar*, 4 mai. 2023. Disponível em: <https://www.musicradar.com/how-to/ai-vocal-covers>. Acesso em: 5 ago. 2023.

HUTCHINSON, C.; JOHN, P. 'Minha obra foi copiada por IA mais do que a de Picasso'. BBC Brasil, 20 jul. 2023. Disponível em: <https://www.bbc.com/portuguese/articles/cqq42vdygqxo>. Acesso em: 5 ago. 2023.

INTELIGÊNCIA artificial da IBM criou 'trailer perfeito' para filme. Olhar Digital, 1 set. 2016. Disponível em: <https://olhardigital.com.br/2016/09/01/pro/inteligencia-artificial-da-ibm-criou-trailer-perfeito-para-filme/>. Acesso em: 5 ago. 2023.

INTERNATIONAL ALLIANCE OF THEATRICAL STAGE EMPLOYEES. 2022 VFX workers survey. 1 mar. 2023. Disponível em: <https://vfxunion.org/2022-survey-results/>. Acesso em: 5 ago. 2023.

JUNQUEIRA, D. DeviantArt rouba arte de usuários para criar ferramenta de IA, mas volta atrás. Olhar Digital, 14 nov. 2023. Disponível em: <https://olhardigital.com.br/2022/11/14/reviews/apos-queixa-de-usuarios-deviantart-altera-sistema-de-arte-criada-por-inteligencia-artificial/>. Acesso em: 5 ago. 2023.

KAHNEMAN, D. *Rápido e devagar: duas formas de pensar*. Rio de Janeiro: Objetiva, 2011.

KALPOKAS, I. Work of art in the age of its AI reproduction. *Philosophy and Social Criticism*, 2023. Forthcoming. DOI: 10.1177/01914537231184490.

KATANICH, D. The world's first AI art gallery opens in Amsterdam. Euronews, 21 mar. 2023. Disponível em: <https://www.euronews.com/culture/2023/03/21/the-worlds-first-ai-art-gallery-opens-in-amsterdam>. Acesso em: 5 ago. 2023.

KEYNES, J. M. Economic possibilities for our grandchildren. 1930. In: KEYNES, J. M. *Essays in persuasion*. New York: Norton, [1931] 1963, p. 358-375. Disponível em: https://archive.org/details/essaysinpersuasiooooojohn_s8g6. Acesso em: 5 ago. 2023.

KRÄMER, S. The artificiality of the human mind: a reflection on natural and artificial intelligence. In: NAGL-DOCEKAL, H.; ZACHARASIEWICZ, W. (eds.). *Artificial intelligence and human enhancement: affirmative and critical approaches in the humanities*. Berlin: De Gruyter, 2022, p. 17-32. DOI: 10.1515/9783110770216-003.

LANDGREBE, J.; SMITH, B. *Why machines will never rule the world: artificial intelligence without fear*. New York: Routledge, 2022.

LOCH, V. C. *A obra de arte na era da inteligência artificial*. 2021. Dissertação (Programa de Pós Graduação em Estudos da Linguagem) β Departamento Acadêmico de Linguagem e Comunicação. Curitiba: Universidade Tecnológica Federal do Paraná, 2021.

MAZZONE, M.; ELGAMMAL, A. Art, creativity, and the potential of artificial intelligence. *Arts*, v. 8, n. 26, 2019. DOI: 10.3390/arts8010026.

MCCORMACK, J. The value of AI art. In: CARVALHAIS, M. et al (eds.) *The Book of X: 10 Years of Computation, Communication, Aesthetics and X*. Porto: i2ADS-Research Institute in Art, Design and Society, 2022, p. 181-192.

MCMAHON, J. Is Hollywood a risky business? A political economic analysis of risk and creativity. *New Political Economy*, v. 24, n. 4, p. 487-509, 2019. DOI: 10.1080/13563467.2018.1460338.

MCMAHON, J. Why Scorsese is right about corporate power. *Class, Race and Corporate Power*, v. 10, n. 1, 2022. Disponível em: <https://www.jstor.org/stable/48658237?seq=3>. Acesso em: 8 ago. 2023.

MENGER, P.-M. Artistic labor markets and careers. *Annual Review of Sociology*, v. 25, p. 541-574, 1999.

- MENGER, P.-M. Artistic labor markets: contingent work, excess supply and occupational risk management. In: GINSBURGH, V. A.; THROSBY, D. (eds.). *Handbook of the economics of arts and culture*. Amsterdam: North Holland, 2006, p. 766-811.
- METCALFE, J. et al. Systemic oversimplification limits the potential for human-AI partnership. *IEEE Access*, v. 9, p. 70242-70260, 2021. DOI: 10.1109/ACCESS.2021.3078298.
- MOURA, F. T.; CASTRUCCI, C.; HINDLEY, C. Artificial intelligence creates art? An experimental investigation of value and creativity perceptions. *Journal of Creative Behavior*, 2023. Forthcoming. DOI: 10.1002/jocb.600.
- MORI, M. The uncanny valley. *IEEE Robotics & Automation Magazine*, v. 19, n. 2, p. 98-100, [1970] 2012.
- MURPHY, G. et al. Face/Off: changing the face of movies with deepfakes. *PLoS One*, v. 18, n. 7, e0287503, 2023. DOI: 10.1371/journal.pone.0287503.
- NATALE, S.; HENRICKSON, L. The Lovelace effect: perceptions of creativity in machines. *New Media & Society*, 2022. Forthcoming. DOI: 10.1177/14614448221077278.
- OLIVEIRA, R. M.; CARVALHO, D. A perspectiva na arte do renascimento. *Educação Gráfica*, v. 18, n. 1, p. 169-182, 2014.
- PINHEIRO, A. M.; TIGRE, P. B. Inovação em serviços. In: RAPINI, M. S. et al (eds.). *Economia da ciência, tecnologia e inovação: fundamentos teóricos e a economia global*. Belo Horizonte: CEDEPLAR, 2021, p. 245-264.
- RAGOT, M.; MARTIN, N.; COJEAN, S. AI-generated vs. Human artworks: a perception bias towards artificial intelligence. CHI EA '20: Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, 2020. DOI: 10.1145/3334480.3382892.
- RAMOS, M. M. Fotografia e arte: demarcando fronteiras. *Contemporânea*, v. 12, n. 1, p. 129-142, 2009. DOI: 10.12957/contemporanea.2009.359.
- ROSCOE, J. AI-generated books of nonsense are all over Amazon's bestseller lists. *Vice*, 28 jun. 2023. Disponível em: <https://www.vice.com/en/article/v7b774/ai-generated-books-of-nonsense-are-all-over-amazons-bestseller-lists>. Acesso em: 5 ago. 2023.
- SALADINO, A. The visual effects crisis. *The Royal Ocean Film Society*, 30 set. 2022. Disponível em: <https://www.andrewsaladino.work/visual-effects-crisis>. 1 vídeo (20:17 min). Acesso em 5 ago. 2023.
- SANCHEZ, T. Examining the text-to-image community of practice: why and how do people prompt generative AIs? C&C '23: Proceedings of the 15th Conference on Creativity and Cognition, jun. 2023, p. 43-61. DOI: 10.1145/3591196.3593051
- SCHOBBER, R. Passing the Turing test? AI generated poetry and posthuman creativity. In: NAGL-DOCEKAL, H.; ZACHARASIEWICZ, W. (eds.). *Artificial intelligence and human enhancement: affirmative and critical approaches in the humanities*. Berlin: De Gruyter, 2022, p. 151-166. DOI: 10.1515/9783110770216-009.
- SHAH, S. The writers strike is taking a stand against AI. *Time*, 4 mai. 2023. Disponível em: <https://time.com/6277158/writers-strike-ai-wga-screenwriting/>. Acesso em: 5 ago. 2023.
- TEIXEIRA, F. O.; MISSIO, F. J. Nível de intensidade tecnológica e lei de Thirlwall multissetorial no

Brasil (1998-2014). Análise Econômica, v. 39, n. 78, p. 145-175, 2021.

THROSBY, D.; ZEDNIK, A. Multiple job-holding and artistic careers: some empirical evidence. *Cultural Trends*, v. 20, n. 1, p. 9-24, 2011. DOI: 10.1080/09548963.2011.540809.

TOKUMITSU, M. *Do what you love: and other lies about success & happiness*. New York: Simon & Schuster, 2015.

TUBADJI, A.; HUANG, H.; WEBBER, D. J. Cultural proximity bias in AI-acceptability: the importance of being human. *Technological Forecasting & Social Change*, n. 173, 2021. DOI: 10.1016/j.techfore.2021.121100.

YINGYING L.; YIXIAO Z. A review on the economics of artificial intelligence. *Journal of Economic Surveys*, v. 35, n. 4, p. 1045-1072, 2021. DOI: 10.1111/joes.12422.

WORLD ECONOMIC FORUM. *Creative disruption: the impact of emerging technologies on the creative economy. White paper*. Geneva: World Economic Forum, 2018. Disponível em: <https://www.weforum.org/whitepapers/creative-disruption-the-impact-of-emerging-technologies-on-the-creative-economy>. Acesso em: 5 ago. 2023.

VALLANCE, C. 'Arte está morta': o polêmico boom de imagens geradas por inteligência artificial. *BBC-News Brasil*, 18 set. 2022. Disponível em: <https://www.bbc.com/portuguese/internacional-62949698>. Acesso em: 5 ago. 2023.

YALALOV, D. *Artstation loses traffic after cracking down on anti-AI protests*. *MetaversePost*, 3 jan. 2023.





■ Giselle Beiguelman, frames do vídeo generativo da série Flora Rebellis, versão 2024.
Imagens geradas com Inteligência Artificial (Style Gans)

■ [Acesse o vídeo clicando na imagem](#)





A EXPANSÃO ARTIFICIAL DA INTELIGÊNCIA HUMANA

DOI: <https://doi.org/10.35699/2965-6931.2023.48125>

LUCIA SANTAELLA*

RESUMO: De poucas décadas para cá, a Inteligência Artificial viu-se transformada em um hype. Não é para menos, pois os efeitos e consequências de suas aplicações em tudo que diz respeito ao humano é indiscutível. Infelizmente, entretanto, quando um tema se torna moeda corrente de publicações nem sempre bem fundamentadas, ele fica submetido a um imbróglio de controvérsias, mal-entendidos e desinformação, justamente o que vem acontecendo com a IA. As razões para isso são muitas. Este artigo pretende colocar em discussão algumas dessas razões, com ênfase no Zeitgeist tanto filosófico quanto tecnocientífico em que a questão da inteligência se acha envolvida.

PALAVRAS-CHAVE: inteligência artificial; hype; Zeitgeist; divulgação científica. abstract The dissemination of art generated by artificial intelligence, or generative art, is changing the creative industries.

The artificial expansion of human intelligence

ABSTRACT: In the last few decades, Artificial Intelligence has become a hype. No wonder, because the effects and consequences of its applications in everything that concerns humans is indisputable. Unfortunately, however, when a topic becomes the currency of publications that are not always well founded, it is subjected to an imbroglio of controversies, misunderstanding and misinformation, precisely what has been happening with AI. The reasons for this state of affairs are many. This article intends to discuss some of these reasons, with emphasis on both the philosophical and the techno-scientific Zeitgeist in which the question of intelligence is involved.

KEYWORDS: artificial intelligence; hype; Zeitgeist; science communication.

* Pontifícia Universidade Católica de São Paulo

Introdução

A inteligência sempre foi um tema de especial interesse da psicologia e também da educação. De umas poucas décadas para cá, entretanto, a inteligência se viu transformada na grande questão do momento. Há três razões fundamentais para isso: a) no campo da filosofia, a emergência do pensamento sobre o não humano; b) na área do meio ambiente, as preocupações e dilemas trazidos pelo Antropoceno, o novo período geológico da Terra e c) o sucesso alcançado pela inteligência artificial que trouxe a inteligência para o centro dos debates.

Este artigo dará breve atenção ao contexto filosófico do não humano e, então, à problemática do Antropoceno, que se constituem em fontes de questionamento das tradicionais concepções do humano particularmente frente às inteligências não humanas. A seguir, atenção mais detalhada será dada ao debate acerca da inteligência, com ênfase na inteligência humana em seu confronto com a inteligência artificial.

A emergência do não humano

Realismo especulativo (RE), virada especulativa ou ainda ontologia orientada ao objeto (OOO) são as denominações do mais recente movimento que marcou um novo caminho para a filosofia. Pode-se determinar seu nascimento e desdobramentos a partir do *Workshop* que ocorreu no *Goldsmiths College*, na Universidade de Londres, em 27 de abril de 2007 e cujo registro consta na íntegra na Revista *Collapse* (Brassier et al., 2007, p. 307-450). Embora tenha nascido na filosofia, seu impacto se fez notar em várias áreas do conhecimento, tais como “antropologia, arqueologia, arquitetura, literatura inglesa, feminismo, artes plásticas, estudos medievais, musicologia, retórica e

composição, estudos científicos e outros” (Harman, 2011, p. vii). Um dos pontos-chave desse movimento encontra-se em sua oposição e mesmo antagonismo à virada linguística que dominou e exerceu grande influência sobre muitos campos do pensamento no século XX. Em seu lugar, o realismo especulativo propôs um outro tipo de virada, ou seja, a virada para o objeto.

São muitos os fatores que poderiam nos levar a explicar por que o RE repercutiu capilarmente em tantas áreas do conhecimento e da produção criativa. Antes de tudo, deve-se considerar o *Zeitgeist* ou sincronias temporais que propiciam que determinados fenômenos, situações ou conjunturas tomem relevo porque aparecem na hora certa e no lugar certo. Assim, sem que houvesse qualquer relação explícita entre ambos, o RE nasceu em meio a debates sobre o pós-humano que já vinham se desenvolvendo nos meios intelectuais desde finais dos anos 1980. Não custa lembrar dos efeitos provocados nos estudos culturais críticos pela publicação do *Manifesto Ciborgue* de Danna Haraway (1985). A tônica do pós-humano versava sobre as transformações do humano provocadas pelas tecnologias. Isso coincidia, sob certos aspectos, com a ruptura defendida pelo RE da relação dual entre sujeito e objeto do conhecimento, sob a primazia do sujeito, vindo daí a virada para o objeto.

O ponto nítido de convergência dos debates do pós-humano com o RE situa-se mais especificamente no deslocamento e descentramento do humano em similaridade de condições com aquilo que Bruno Latour (1994a, p. 140; 1994b) chamou de “parlamento das coisas”, no qual os novos protagonistas são os actantes, tudo que nos rodeia e que utilizamos em nossas práticas, inclusive científicas, a saber, objetos, ferramentas, instrumentos que “fazem fazer” de acordo com as operações de tradução, mediação, intermediários, mediadores etc., operações nas quais não têm mais lugar as relações dicotômicas pautadas na primazia do sujeito do conhecimento ou do fazer sobre a subserviência do objeto (ver Lemos, 2013).

O parlamento das coisas brotou na consonância com a Teoria Ator-Rede (TAR) e seu ataque agudo às categorias e conceitos opositivos que dominaram o cenário do pensamento ocidental por séculos. Para a TAR, verdade/falsidade; grande/pequeno; agência/estrutura; humano/não humano; antes/depois; conhecimento/poder; contexto/conteúdo; materialidade/socialidade; atividade/passividade são visões separatistas que foram borradas no trabalho realizado pela TAR. A associação do pensamento de Latour com a TAR foi inevitável. Na sua versão, pessoas, animais, coisas, objetos, apa-

relhos, dispositivos e instituições podem funcionar como actantes em rede, ou seja, agenciadores que fazem agir (Latour, 1996, 1998; Santaella e Lemos, 2010). Não obstante a reviravolta que provoca nos esquemas conservadores de interpretação do mundo, as propostas da TAR expandiram-se da sociologia da ciência e da tecnologia para a sociologia em geral, a geografia, a administração, os estudos das organizações, a antropologia e a filosofia.

Ademais, a parte fundamental do *Zeitgeist* em sintonia com a virada para o objeto do RE, conforme foi discutido em Santaella (2013), encontra-se no advento da internet das coisas (IoT) que está transformando em seres sencientes, comunicacionais, aquilo que, com o nome de meras “coisas”, costumávamos tratar como seres inertes ao nosso dispor. Essa emergência, no contexto de outras invenções que a digitalização ubíqua está trazendo, reclama pela renovação do pensamento ontológico, especialmente da ontologia do humano que herdamos do passado.

Não levou muito tempo para que tanto o RE quanto a TAR fossem absorvidos, desde o início da segunda década do nosso século, no contexto maior que ficou conhecido como a virada do não humano (nonhuman turn) que engloba estudos interdisciplinares das mais diversas ordens, todos eles na defesa do descentramento do humano no seio da biosfera. Assim, o não humano passou a englobar o mundo animal e seus direitos, a teoria dos afetos mobilizada pelas discussões sobre diversidade sexual, as interrogações sobre os corpos, os sistemas orgânicos e geofísicos, as novas materialidades nas teorias feministas, as neurociências, as ciências cognitivas, as vidas artificiais, as questões da inteligência e consciência. Em suma, um pool de complexidades que desafiam nossas compreensões da realidade e que exigem o nosso engajamento com o que não é humano, tais como mudanças climáticas, secas, fome, biotecnologia, genocídio, terrorismo, guerra e até mesmo o Antropoceno, o novo período geológico do planeta, fruto do peso e feridas que as ações humanas, muitas vezes insanas, imprimiram sobre a biosfera (GRUSIN, 2015, p. vii).

O antropoceno e suas consequências

Foi só em 2000 que a denominação Antropoceno, para marcar uma mudança no período geológico da Terra, foi proposta por Paul Z. Crutzen, e por Eugene F. Stoemer, em uma publicação do *International Geosphere-Biosphere Programme*. Sem nos determos em detalhes científicos complexos, o Holoceno, no qual estávamos inseridos, é uma época geológica pós-glacial, cujo início data de doze mil anos. Trata-se de uma época marcada, por exemplo, pelo advento da agricultura, então, pela formação de cidades, meios de comunicação e por novas formas de energia. Durante toda essa época, o regime climático estabilizou-se com climas temperados e os humanos se fizeram presentes em todos os continentes.

Entretanto, nas últimas décadas, foi se criando um consenso entre os cientistas de que a biosfera deixou de se comportar dentro dos padrões do Holoceno: a atmosfera, o clima, os oceanos e o ecossistema estão todos operando fora das normas holocênicas. A antropoesfera, a camada humana, que cresceu dentro da biosfera, deixou pegadas destrutivas profundas sobre a Terra a ponto de dar passagem a um novo período geológico, o Antropoceno (do grego *anthropos* = ser humano + *Kainos* = recente, novo).

Há muitas discussões relativas ao ponto no tempo em que esse novo período geológico teve início. Duas opções dominam nas discussões: final do século XVIII, com a revolução industrial, ou então 1945, depois da Segunda Guerra Mundial. No primeiro caso, a análise do ar mostra o início da concentração global crescente do dióxido de carbono e metano, o que coincide com a invenção da máquina a vapor, 1784, por James Watt. É a partir dessa data também que teve início a explosão demográfica que então estava na margem de um bilhão, enquanto existem prognósticos de que neste século a população chegue a dez bilhões de pessoas. As florestas tropicais desaparecem em um ritmo alarmante soltando dióxido de carbono e levando à extinção de várias espécies.

No segundo caso, diante da “grande aceleração”, ou melhor, da mudança do ritmo da própria mudança instaurada a partir do final da Segunda Guerra, 1945 é sinalizado como data inicial do Antropoceno. Foi essa época que marcou o surgimento da era nuclear e da assinatura radioativa que ela deixou sobre a superfície da Terra. Paralela-

mente, a economia global cresceu dez vezes em apenas algumas décadas, provocando a perda da biodiversidade, com o ritmo de extinção das espécies da biosfera correndo cem a mil vezes mais rapidamente. Enfim, o grande problema é que, acompanhando esse ritmo, estamos consumindo exponencialmente mais recursos da natureza.

Até recentemente, o estudo das eras, épocas e períodos geológicos parecia importar pouco para o restante dos campos de conhecimento e, certamente, importava menos para os leigos. Entretanto, se a revolução copernicana já foi capaz de transformar a visão humana sobre o ambiente em que vive, agora não há como deixar de compreender criticamente como o ser humano está implicado nos destinos do planeta. Diante de tudo isso, o tema do Antropoceno, nos últimos anos, deixou de ser privilégio dos especialistas, espalhando-se por várias áreas do conhecimento, inclusive das artes, e repercutindo cada vez mais intensamente nas mídias informativas, tais como jornais, revistas, TV e internet.

Relevante para a questão da inteligência humana e não humana, é notável que as condições do Antropoceno provocaram a emergência de movimentos ativistas e artistas a par de fértil produção discursiva que alertam para uma necessária transformação da noção exclusivista da inteligência humana, demonstrando em pesquisas legitimadas a existência de muitas formas de inteligência não humanas que se estendem por todo o mundo vegetal, naquilo que ficou conhecido como a revolução das plantas (Mancuso, 2019) até as mais diversas modalidades de inteligência animal (Bridle, 2023). Tanto o RE e a TAR quanto o Antropoceno nas suas reivindicações, que são também de natureza política, proclamam a necessidade de expandir a noção de inteligência para além de sua reclusão estrita no reino do humano. É nesse contexto que as versões bem-sucedidas da Inteligência Artificial (IA) emergiram, trazendo, elas próprias novos questionamentos sobre a inteligência em geral, a inteligência humana, a inteligência artificial e a relação entre ambas.

A emergência bem-sucedida da inteligência artificial

Desde seu nascimento em meados do século passado, a inteligência artificial é um campo de conhecimento que pertence ao feixe das ciências cognitivas, um conjunto de ciências distintas, mas que se unem no foco comum das pesquisas sobre as habilidades cognitivas capazes de elucidar os aspectos constitutivos da inteligência. Nos seus setenta anos de existência, a IA passou por momentos de euforia, seguidos por disforias que só se dissiparam há perto de duas décadas.

De fato, atualmente a Inteligência Artificial (IA) se constitui no *hype* das pesquisas tecno científicas com repercussões em muitas esferas das atividades humanas. Entretanto, quando se trata de definir a IA as dificuldades surgem, especialmente porque a presença do substantivo “inteligência” só traz dificuldades, pois são muitas as definições de inteligência nem sempre concordantes. Quando se acrescenta o adjetivo “artificial”, as dificuldades só tendem a aumentar. Diante disso, é preciso buscar um substrato de concordância que nos leve a uma definição breve e relativamente consensual extraída daquilo que é comum em vários textos sobre o tema.

Assim, a IA significa a simulação pelos sistemas computacionais de processos de inteligência humana. Trata-se de um ramo das ciências da computação voltado para a criação de máquinas inteligentes. Isso implica o desenvolvimento maquínico de habilidades tais como são, por exemplo, a aprendizagem, o conhecimento, a aquisição de informação e as regras para usá-la, o raciocínio utilizado para alcançar conclusões definidas ou aproximadas, a autocorreção, a solução de problemas, a percepção, o reconhecimento e processamento linguístico, o planejamento, e a habilidade para manipular e mover objetos. Para realizar esses fins, o computador precisa ter acesso a objetos, categorias, propriedades e relações.

De uns anos para cá, a euforia provocada pela IA é fruto de vários fatores. Entre eles, principalmente o aumento considerável do poder computacional, o agigantamento dos dados digitais e as redes neurais artificiais. Desde que as válvulas elétricas foram substituídas pelos transistores (1947), as máquinas computacionais, munidas de microprocessadores (1971), foram diminuindo de tamanho proporcionalmente ao

aumento de seu desempenho. A escalabilidade ou capacidade de *upgrade* das máquinas acompanhou o crescimento progressivo de dados a serem processados. Há indicadores não atualizados de que as informações do mundo dobram exponencialmente a cada dois anos, sendo referidas como *Big Data*, caracterizado não apenas por seu volume, mas por uma rica mistura de tipos e formatos de dados (variedade) e pela natureza sensível ao mesmo tempo que marca um desvio do processamento em lote tradicional (velocidade). Esse enorme volume de dados permitiu o funcionamento mais eficaz das redes neurais artificiais cujo surgimento não é tão recente como se pode pensar.

Na história das ciências cognitivas, as redes neurais surgiram como reação à primeira fase da inteligência artificial regida pelo modelo computacional da mente. Conforme já desenvolvido em Santaella (2004, p. 77-78), dos anos 1950 em diante, a inteligência artificial passou a buscar as condições formais da atividade cognitiva, capazes de indicar o que é comum a todos os sistemas que exibem essa atividade, quer ela apareça em animais, máquinas ou humanos. Para isso, o modelo computacional apresentava habilidades para simular processos cognitivos, com promessas de modelar a mente. Em vista disso, esse modelo dominou por certo tempo como o paradigma clássico unificador das ciências cognitivas.

O modelo partia de dois pressupostos: antes de tudo, “a relativa autonomia entre o *hardware* e o *software* das máquinas utilizadas para simular a inteligência, o que permite explicar o comportamento inteligente de qualquer sistema complexo sem pressupor o tipo físico ou biológico da inteligência de seus componentes”. Isso conduzia à “compreensão da mente como um conjunto de representações de tipo simbólico e regidas por um conjunto de regras sintáticas” (Santaella, *ibid.*). Conclui-se, a partir disso, que o pensamento seria o resultado da ordenação mecânica de uma série de representações ou símbolos, ordenação esta que não pressupõe necessariamente a existência de um cérebro. Por isso, o aparato mental concebido como dispositivo lógico pode ser descrito por meio de um conjunto de computações abstratas. Nessa perspectiva, simular a inteligência não implica a construção de máquinas com *hardware* específicos, mas sim o desenvolvimento de programas computacionais operando sobre dados ou representações. Chamado de inteligência artificial simbólica, esse paradigma sedimentou-se a partir de final dos anos 60 tanto nos trabalhos do grupo liderado por Newell e Simon quanto nos de Marvin Minsky e Seymour Papert (Teixeira, 1998, p. 36, 43).

Não é difícil inferir que está aí implicada uma mudança no conceito de inteligência que passa a ser definida como capacidade para produzir e manipular símbolos, tendo em vista a resolução de problemas. Essa concepção da inteligência e da mente teve e continua tendo consequências filosóficas que não são desprezíveis, especialmente para as seculares questões sobre as relações entre mente e corpo e, em tempos mais recentes, diante dos avanços das neurociências, para as relações entre mente e cérebro. A teoria da mente que está por trás do paradigma computacional é o funcionalismo que exerceu domínio quase exclusivo nas ciências cognitivas até os anos 1980 quando se deu uma virada para o conexionismo. Aproveitando-se dos avanços da época das pesquisas sobre redes de neurônios formais, o conexionismo propôs a replicação da inteligência por meio da construção de redes neurais artificiais. Usando técnicas dotadas de propriedades que podem ser interpretadas em termos cognitivos, essas redes são capazes de aprender, reconhecer formas, memorizar por associações etc.

Enquanto o modelo computacional da mente, também chamado de cognitivismo, estuda os processos mentais como computações abstratas, independentemente de suas formas específicas de concretização, o conexionismo pretende simular o cérebro como meio para simular a atividade mental. Os conexionistas tomam o cérebro humano como um dispositivo computacional em paralelo que opera com milhões de unidades similares aos neurônios. Se computadores e cérebros têm como função principal processar informações, então redes neurais artificiais podem ser construídas para simular esse processo.

As redes neurais constituem uma intrincada rede de conexões entre suas unidades que são dispostas em camadas hierarquicamente organizadas. Conectadas umas com as outras, unidades estimuladas via *inputs* externos excitam ou inibem outras unidades gerando padrões de conectividade. Diferentemente dos sistemas computacionais simbólicos, os conexionistas são sistemas dinâmicos compostos por um “conjunto de processos causais através dos quais as unidades se excitam ou se inibem, sem empregar símbolos e tampouco regras para manipulá-los”. Pretende-se que esse conjunto de neurônios artificiais modele a cognição (Teixeira, 1998, p. 84). Embora infinitamente menos complexo do que o cérebro, esse modelo da mente inspira-se na estrutura e modo de funcionamento do cérebro, chegando assim mais próximo da realidade biológica da mente. O que importa em sistemas desse tipo são os complexos padrões de

atividade entre as múltiplas unidades que constituem uma rede (ibid., p. 103).

O conexionismo representou um passo importante para a compreensão do funcionamento da IA. Entretanto, o tempo não estava maduro para o seu desabrochar. O poder computacional ainda estava longe de alcançar o patamar em que se encontra desde as duas últimas décadas. Então, a quantidade dos dados era incipiente. Nos anos 1980, a WWW, as redes sociais e outras formas de captura de dados não haviam acontecido. Era preciso que tudo isso viesse à tona para que a IA emergisse com a força que hoje tem e, graças à grande estrela da aprendizagem de máquina, transformar-se no hype do momento, a ponto de asfixiar muitos outros.

A ênfase na inteligência

A ênfase na questão da inteligência está colocada, desde meados do século passado, no âmbito das ciências cognitivas de que as ciências da computação e hoje a ciência dos dados, acrescidas das pesquisas em IA, são partes constitutivas. A par disso, para compor o *Zeitgeist*, os especialistas em tecnologias em geral e na filosofia da tecnologia em particular integraram o tema da inteligência em seus estudos, com atenção para as transformações pelas quais o conceito de inteligência vem passando no contexto de uma preocupação com a necessidade de se repensar a própria ontologia do humano.

Afinal, a espécie humana é uma espécie em evolução pelo simples fato de que o *Sapiens* é uma espécie que se constitui como inalienavelmente inacabada, sempre em busca de si mesma e de um alvo que, embora sejamos parte dele, desconhecemos. Como se não bastasse, tanto a filosofia que ilumina as mutações sociotécnicas próprias da digitalização e contemporaneamente da dataficação e performatividade algorítmica, à luz da TAR, quanto também os debates sobre o Antropoceno vêm proclamando a urgência do deslocamento e descentramento do humano junto com o conceito de inteligência com que se viu acompanhado pela tradição.

No contexto desse *Zeitgeist* é curioso observar como o campo da IA -- especialmente o campo daqueles que rodeiam a IA e recebem seus efeitos, mas não estão den-

tro da especialidade das pesquisas -- tem revelado uma resistência diante da premência do enfrentamento do fato de que a IA está exigindo uma transformação na própria ideia daquilo que o passado nos fez acreditar sobre o que é inteligência, a saber, no caso, como sendo uma exclusividade humana intocável.

Mais curioso ainda é notar que a proposta de expansão da noção de inteligência para o compartilhamento da inteligência humana com aquelas dos animais e mesmo das plantas é recebida sem dificuldades, enquanto o compartilhamento com a inteligência das máquinas, que atualmente alcançou o estado da arte da IA generativa, é recebido com os mais diferentes tipos de entraves. Como explicar essa dificuldade? A tradição nos mostra que a busca de separação entre o orgânico, particularmente o organismo humano, e as máquinas vem de longe e se encontra bem representada tanto na filosofia cartesiana como em versão mais renovada na filosofia de Heidegger.

Sugestões de resposta para as oposições entre humano e máquinas podem ser encontradas no fato de que o humano pode experimentar muitos afetos, inclusive paixão e piedade; máquinas não sentem nada. Elas funcionam dentro de limites estreitamente definidos, e quanto mais padronizados, mais esses limites se tornam rígidos. Os organismos, ao contrário, são mais abertos às potencialidades do acaso, das surpresas, das renovações e mesmo das mutações. A vida é experiência, o que significa improvisação, a vida é tentativa e busca em todos os respeitos.

Na concepção obsoleta de máquinas que ainda abrigamos, máquinas são compostas de partes. Elas são armadas e desarmadas. Elas estão abertas a modificações ou reconstruções. Para o ser humano, a ideia de ganhar novas partes é apenas um pouco menos horrível do que perder partes. A máquina tanto solta quanto adquire novas partes com a maior facilidade. A perspectiva humana parece insistir sobre a integridade orgânica como a única norma possível (Wilson 1996, p. 246). Tudo isso indica que “somos vítimas de uma esperança nostálgica de retorno à plenitude de um corpo uno-primordial. Uma esperança teimosamente renovada, mas fadada ao malogro, pois tal plenitude só existe na saudade sonhada da carícia materna que tremeluz nas neblinas da imaginação” (Santaella, 2003, p. 228).

Isso explica por que, dentre todas as linguagens, técnicas e tecnologias, só a fala e o gesto são sentidos como naturais enquanto as outras técnicas são tidas como artificiais, estranhas, irreconhecíveis. Isso ocorre porque, não obstante sejam também

técnicas, fala e gesto ainda estão inseridos no próprio corpo, são ainda partes integrantes do corpo e, portanto, não são percebidas como técnicas. O estranhamento e as resistências manifestam-se quando as técnicas começam a se lançar para fora do corpo como extrojeções de habilidades do cérebro, a extrojeção da memória, por exemplo, na escrita, ou extrojeções do gesto, na máquina fotográfica como amplificação do sentido da visão, ou nos aparelhos eletroacústicos como amplificação do sentido da audição. Desde essas extensões sensoriais até a emergência da cultura mediada pelo computador, as máquinas deixaram de funcionar como meras ferramentas, para se transformarem em tecnologias da inteligência em evolução acelerada que, no ponto em que estamos, alcançaram a simulação das faculdades discursivas e imagéticas humanas. É o estranhamento e mesmo preconceito diante das tecnologias inteligentes, que a IA leva a um certo limite, que se constitui na celeuma em torno do tema da inteligência que hoje reina no campo da IA, mesmo antes do abalo sísmico que a IA generativa vem provocando.

O imbróglio da inteligência

No seu estágio atual a IA é movida por aprendizado de máquina (AM) e por sua vertente mais bem sucedida, a aprendizagem profunda (AP). De acordo com a literatura especializada, existem pelo menos cinco grupos diferenciados de desenvolvimento da IA: os simbolistas, as conexionistas, os evolucionários, os bayesianos e os analogistas. Tudo indica que, entre eles, o grupo que tem alcançado maior sucesso, pela performatividade de suas aplicações, é o conexionismo.

Conforme já desenvolvido em Santaella (2023), na sua versão da IA preditiva, antes da IA generativa, o conexionismo opera com redes neurais artificiais mais profundas do que em qualquer outra época e elas definem novos padrões para a visão, o reconhecimento da voz, a descoberta de medicamentos e outras áreas. Para os conexionistas, aprendizado é a função do cérebro que aprende ao ajustar as forças das conexões entre neurônios, e o problema crucial é descobrir quais conexões são responsáveis por quais erros e alterá-las de acordo. Mimeticamente ao funcionamento das redes sinápticas

do cérebro, o algoritmo mestre dos conexionistas é a *backpropagation* (propagação retrógrada, retropropagação ou propagação reversa), que compara a saída de um sistema com a saída desejada e altera sucessivamente as conexões em cada camada de neurônios para aproximar mais a saída daquilo que ela deveria ser, ou melhor, que se espera que ela seja (Domingos, 2017, p. 77-78).

Esses rudimentos de explicação técnica parecem suficientes para colocar em discussão aquilo que importa neste artigo, ou seja, a inteligência humana e a IA, colocando em relevo o imbróglio em que esse tema está hoje envolvido. Como ponto de partida, é preciso considerar que a IA é uma área de pesquisa ultra especializada, envolvendo complexidades matemáticas, estatísticas e probabilidades que não estão ao alcance dos leigos. Todavia, quando um *hype* toma conta da cultura, da política e da economia, todos se julgam autorizados a se pronunciar sobre o tema, mesmo quando não tiveram a menor curiosidade para enfrentar as informações especializadas, tanto internas quanto externas sobre a questão. Esse é o péssimo legado que as redes sociais nos deixaram e que se alia ao costumeiro sensacionalismo das mídias noticiosas que nadam a braçadas naquilo que pode atrair seu público. A par disso, não obstante a indiscutível relevância da divulgação científica, ela pode produzir um desserviço quando se deixa contaminar pelo sensacionalismo ou pela autocomplacência.

O primeiro grande foco do imbróglio, composto por controvérsias, equívocos e desinformação vem da confusão entre consciência e inteligência. Felizmente os pesquisadores estão longe de cair nessa cilada, antes de tudo porque sabem que seu *metier* diz respeito à inteligência e não à consciência, esta última uma questão que é melhor ser deixada para os filósofos e psicólogos e, daqui a algum tempo, provavelmente para a neurociência.

De fato, há um certo consenso entre os pesquisadores de IA de que consciência não se confunde com inteligência, embora, no animal humano, ambas estejam indiscutivelmente entrelaçadas. Infelizmente na filosofia cognitiva, a consciência tem sido reduzida às discussões sobre o *qualia* quer dizer, “o sentido subjetivo de uma experiência difícil de capturar em quaisquer tipos de sistemas formais”: por exemplo, a vermelhidão do vermelho como é sentida por cada um de nós. “Tomamos como garantido que a outra pessoa está vendo a mesma cor, mas não sabemos se ela tem a mesma experiência subjetiva que eu tenho.” O se-

gundo aspecto refere-se ao sentido do self. Experimentamos o mundo “em uma certa espécie de modo unitário e experimentamos a nós mesmos como estando nele.” Não o experimentamos em termos de dezenas de milhões de neurônios acesos (Tenenbaum, 2018, p. 482-483).

Outros pontos de convergência dos especialistas acerca da consciência dizem respeito à persistência no tempo da autoconsciência (Bostron (2018, p. 11), a saber, à constatação de que a consciência consiste em nossas experiência subjetivas, ou seja, consiste em nosso mundo interior intransferível e intraduzível. Diante disso, Koller (2018, p. 395) afirma que chegar à IA Geral ou superinteligência dispensa a consciência, “pois é possível ter um sistema incrivelmente inteligente, que não tem nada a ver com uma consciência interior.” Turing já havia afirmado que a consciência não é conhecível. Tanto isso é verdade que o surgimento da IA generativa tem sido acompanhada pelos prognósticos de uma possível inteligência geral para o futuro, nos quais o tema em foco é a inteligência em si e não a consciência.

De todo modo, a partir dessas convergências, tendo a concordar que a consciência é um fenômeno interior, e por ser interior, é difícil encontrar formas de compartilhamento entre consciências, tanto quanto é difícil saber se outros animais também a têm. Por outro lado, por ser simultaneamente interior e exterior, a inteligência humana, por seu lado, é compartilhável por meio de linguagens e sistemas de codificação externos à mente, tendo, por isso, uma natureza coletiva (para mais detalhes, ver Santaella, 2022 e 2023, onde essa questão é tratada detalhadamente).

O outro foco do imbróglio em que a questão da inteligência, em especial a computacional está envolvida, provém dos negacionistas da inteligência artificial, quer dizer, da negação peremptória de que haja qualquer forma de inteligência na inteligência artificial, o que leva à consideração de que o substantivo “inteligência” não lhe cabe. Essa negação se constitui em uma verdadeira ladainha de vozes.

Na maior parte das vezes, o que está por trás dessa ladainha é o *parti-pris* de que inteligência é tão só e apenas humana. Quaisquer outras formas de inteligência são, portanto, corruptelas de inteligência ou nem mesmo inteligência, um problema que é criticamente comentado por Rus (2018, p. 263). Para essa autora, “há um grande equívoco na imprensa popular sobre o que é e o que não é IA. Quando dizem IA referem-se a aprendizado de máquina e, mais que isso, aprendizagem profunda dentro da

aprendizagem de máquina. As pessoas tendem a antropomorfizar o que esses termos significam. Os não especialistas associam com inteligência e tomam esta como inteligência humana.

Quanto ao termo “artificial”, ele é utilizado para se referir à inteligência das máquinas e não à expressão mais ampla de inteligência não humana que inclui a inteligência dos animais para aqueles que não se limitam a uma visão antropocêntrica da inteligência. Para Russel (2018, p. 63) uma entidade pode ser considerada inteligente na medida em que faz a coisa certa, significando que se espera que suas ações atinjam seus objetivos. Esse princípio aplica-se tanto aos humanos quanto às máquinas. “Fazer a coisa certa é a chave do princípio unificador da IA. Para isso, a IA precisa de algumas habilidades-chave, percepção, visão, reconhecimento de fala, e ação. Especificando, trata-se “da habilidade de tomar decisões, planejar, e resolver problemas, também da habilidade de comunicar, portanto, o entendimento da linguagem natural se torna muito importante para a AI” (ibid.).

Um argumento que abre caminhos contra os discursos negacionistas baseia-se na questão do aprendizado. A inteligência humana é viva e metabólica, implicando o aprendizado contínuo, base para a consideração da plasticidade do cérebro. Mas, se o aprendizado se encontra no coração das habilidades de que a inteligência é feita, esse parece ser o nó-górdio que implica considerar equivocado negar a inteligência às máquinas. Sim, os algoritmos de IA são inteligentes e, assim o são, primordialmente porque aprendem, auto-organizam-se, corrigem seus erros e chegam aos seus alvos. Tanto é assim que a comparação entre os processos humanos de aprendizagem e aqueles dos algoritmos tornou-se um tema de investigação dos especialistas em IA, isto porque, certamente, para construir máquinas inteligentes, eles não podem escapar da comparação com a complexidade dos humanos.

De fato, os defensores da inteligência computacional consideram a IA inteligente porque o computador adquiriu o potencial de aprender e tomar decisões com base nas informações que recebe. Isso não desconsidera o fato de que esse tipo de inteligência é decididamente diferente da nossa. Por suas habilidades de imitação, a IA tem a qualidade de identificar padrões informacionais que otimizam tendências relevantes para os resultados a que chegam. Além disso, ao contrário dos humanos, a IA não possui um corpo metabólico regulado por hormônios, um corpo que sente fadiga, tem sono,

um corpo inseparável da mente que é muitas vezes atormentada por complicações psicológicas que, inclusive, podem ser somatizadas. Por ser corporificada a inteligência humana está submetida às flutuações químicas do nosso corpo, flutuações, de resto, que têm propósitos. A nossa máquina química interfere e, por vezes, supera a máquina raciocinante diante de urgências, pois as emoções são uma máquina química de ajustamento de nossas prioridades existenciais.

Para compensar as exigências corporais, as habilidades humanas ganham em expansividade, pois, enquanto a IA só responde aos dados disponíveis, os humanos têm a capacidade de imaginar, antecipar, sentir e julgar situações de mudança. Nós humanos somos providos de sentimentos e afetos, inclusive tristes, que nem nós mesmos capazes de compreender. Nas sábias palavras de Fernando Pessoa, “amamos sempre no que temos o que não temos quando amamos”, fonte primeira de nossas contradições. Somos dotados de senso comum, pois vivemos em um mundo de experiências factuais que muitas vezes nos dão força e outras nos fragilizam.

Em função de tudo isso, não fazem sentido as comparações apressadas entre a inteligência humana e a IA, nem mesmo no ponto em que nos encontramos da IA generativa que aprendeu a conversar conosco. Mas igualmente não faz sentido afirmar que a IA não tem inteligência pelo simples fato de que suas modalidades de inteligência são distintas da nossa. Embora distintas, a IA opera eficazmente com o raciocínio dedutivo, uma forma de inteligência que também é processada por mentes humanas, com a diferença de que a IA é capaz de lidar com um grande número de dados com os quais as redes neurais artificiais aprenderam a lidar.

Na verdade, do confronto entre IA e inteligência humana, origina-se um paradoxo: o que é difícil para o humano, a IA faz, o que é difícil para a IA, o humano faz. Na realidade são dois tipos distintos de inteligência e, por serem distintos, por mais que a inteligência humana possa ou não nos orgulhar, isso não permite que se negue a potencialidade cognitiva e a expressão da inteligência na IA sem negar que há características e propriedades que funcionam como entraves para que ambas possam se igualar.

O que se pode concluir com Lecun (2018, p. 133) é que os sistemas de IA irão amplificar a inteligência humana do mesmo modo que as máquinas mecânicas foram uma amplificação da força física. Não serão uma substituição. Em suma, devemos prever um futuro em que sistemas computacionais terão habilidades complementares

às humanas, funcionando muito mais como uma inteligência aumentada na união de dois tipos de inteligência distintos.

O recrudescimento dos princípios iluministas

Desafortunadamente, pouco efeito podem ter os argumentos a favor do tipo de inteligência de que a IA está dotada frente à exacerbação anacrônica a que estamos assistindo dos princípios iluministas como capa de proteção da exclusividade da inteligência humana contra quaisquer modalidades de IA.

Tenho repetido com certa insistência que o advento da IA trouxe consigo um infeliz revigoração de velhos e cínicos valores iluministas (ver Sloterdijk, ...) , agora anacronicamente tidos como o suprasumo do humano. A filosofia do século XX demoliu esses valores. Essa demolição teve continuidade no pós-humano crítico e na teoria do ator-rede entre outras, assim como na necessidade imperiosa de reconsideração da ontologia do humano frente à sua solidariedade com outras formas de inteligência da biosfera. De onde vem tanto temor e protecionismo contra formas de inteligência não-orgânicas, processadas no interior de tecnologias inteligentes?

Se continuarmos a pensar a relação entre humano e IA, de modo opositivo, a partir de valores herdados do Iluminismo (sujeito, individualismo, intencionalidade, livre-arbítrio etc.), ou seja, valores que engordam o egocentrismo e que foram demolidos pela filosofia da desconfiança (Foucault, Deleuze, Derrida, Lyotard, que infelizmente tão poucos leram), nossa compreensão da IA ficará presa ao imbróglio sensacionalista (para o bem da mídia) em que nos encontramos.

Levar em consideração as modalidades de inteligência operadas pela IA não significa ocultar ou negligenciar os imensos desafios que a IA está trazendo para tudo o que diz respeito ao humano. O caráter fundamental das tecnologias, desde sempre, encontra-se na ambivalência. Se, de um lado, a IA e suas aplicações trazem, de fato, contribuições inestimáveis para a gestão da vida humana, de outro lado, não devem ser minimizadas as contradições e dilemas que ela traz e que têm sido objeto de discussões éticas levadas a efeito no campo das regulamentações cuja urgência não pode ser pro-

crastinada. Contudo, contradições e dilemas devem ser enfrentados cara a cara. Disfarces e dissimulações proferidos em nome de um protecionismo baseado em concepções obsoletas do humano só colaboram para o afastamento dos verdadeiros problemas.

Referências

BOSTROM, Nick. Entrevista concedida a Martin Ford. In Ford, Martin (ed.). *Architects of intelligence. The truth about AI from the people building it*. Birmingham: Packt Publishing, p. 97-116, 2018.

BRIDLE, James. *Maneiras de ser. A busca por uma inteligência planetária*. Trad. Daniel Galera. São Paulo: Todavia, 2023.

BRYANT, Levi; SRNICEK, Nick; HARMAN, Graham. *Toward a Speculative Philosophy*. In *The Speculative Turn: Continental Materialism and Realism*, Levi Bryant, Nick Srnicek, and Graham Harman (eds.). Melbourne: re.press, 2011.

CRUTZEN, Paul; STOERMER, Eugene F. The “anthropocene”. *IGBP Newsletter*, n. 41, p. 17-18, May 2000. Disponível em: <http://www.igbp.net/download/18.316f18321323470177580001401/NL41.pdf>. Acesso: 10 ab, 2022.

GRAHAM, Harman. *Quentin Meillassoux. Philosophy in the making*. Edinburgh University Press. 2011.

GRUSIN, Richard. Introduction. In Grusin, R. (ed.). *The nonhuman turn*. Minneapolis: The University of Minnesota Press, p. vii-xxix, 2015.

HARAWAY, Donna. Manifesto ciborgue: ciência, tecnologia e feminismo-socialista no final do século XX. In Silva, Tomaz Tadeu (org, trad.). *Antropologia do ciborgue*. Belo Horizonte: Autêntica, p. 37-130, (1985) 2000.

KOLLER, Daphne. Entrevista concedida a Martin Ford. In Ford, Martin (ed.) *Architects of intelligence. The truth about AI from the people building it*. Birmingham: Packt Publishing, p. 387-403, 2018.

LATOUR, Bruno. *Jamais fomos modernos. Ensaio de antropologia simétrica*. Trad. de Carlos Irineu da Costa. Rio de Janeiro: Ed. 34, 1994a.

_____. *On technical mediation: Philosophy, sociology, genealogy*. *Common Knowledge*, 3 (2), p. 29-64, 1994b.

_____. *On actor-network theory: a few clarifications*. *Soziale Welt*, p. 369-381, 1996.

_____. *On recalling ANT*. In Law, J; Hassard, J. (eds.) *Actor-network theory and after*. Oxford: Blackwell Publishers, p. 15-26, 1998.

LECUN, Yann. *Entrevista concedida a Martin Ford*. In Ford, Martin (ed.). *Architects of intelligence. The truth about AI from the people building it*. Birmingham: Packt Publishing, p. 119-142, 2018.

LEMOS, André. *A comunicação das coisas. Teoria ator-rede e cibercultura*. São Paulo: Annablume, 2013.

MANCUSO, Stefano. *A revolução das plantas*. São Paulo: Ubu, 2019.

RUS, Daniela. *Entrevista concedida a Martin Ford*. In Ford, Martin (ed.). *Architects of intelligence. The truth about AI from the people building it*. Birmingham: Packt Publishing, p. 253-268, 2018.

RUSSEL, Stuart J. *Entrevista concedida a Martin Ford*. In Ford, Martin (ed.) *Architects of intelligence. The truth about AI from the people building it*. Birmingham: Packt Publishing, p. 39-68, 2018.

SANTAELLA, Lucia. *Pressupostos teóricos da cognição*. In Santaella, Lucia. *Navegar no ciberespaço. O perfil cognitivo do leitor imersivo*. São Paulo: Paulus, p. 73-92, 2004.

_____. *O pós-humano na convivência com a internet das coisas*. In Santaella, Lucia. *Comunicação ubíqua. Repercussões na cultura e na educação*. São Paulo: Paulus, p. 23-38, 2013.

_____. *Neo humano. A sétima revolução cognitiva do Sapiens*. São Paulo: Paulus, 2022.

_____. *A inteligência artificial é inteligente?* São Paulo: Almedina. Edições 70, 2023.

SANTAELLA, Lucia e Lemos, Renata. *Redes sociais digitais. A cognição conectiva do Twitter*. São Paulo: Paulus, 2010.

TEIXEIRA, João de Fernandes. *Mentes e máquinas. Uma introdução à ciência cognitiva*. Porto Alegre: Artes Médicas, 1998.

TENENBAUM, Joshua. *Entrevista concedida a Martin Ford*. In Ford, Martin (ed.). *Architects of intelligence. The truth about AI from the people building it*. Birmingham: Packt Publishing, p. 473-491, 2018.



■ Giselle Beiguelman, frames do vídeo generativo da série Flora Rebellis, versão 2024.
Imagens geradas com Inteligência Artificial (Style Gans)

■ [Acesse o vídeo clicando na imagem](#)





OBSTÁCULOS PARA UMA MOBILIDADE URBANA AUTÔNOMA A PARTIR DA PERSPECTIVA HUMANA: Uma revisão sistemática

DOI: <https://doi.org/10.35699/2965-6931.2023.46814>

SANDOR BANYAI PEREIRA*; IARA SOUSA CASTRO**; RÓBER DIAS BOTELHO***

RESUMO: Ao longo de séculos de desenvolvimento, foi possível ao ser humano criar e desenvolver novas ferramentas que suprimissem necessidades de mobilidade dentro e fora do ambiente urbano. Diversas tecnologias estão sendo convergidas para um novo tipo de mobilidade: a Mobilidade 4.0. O presente artigo objetiva discutir os principais obstáculos para a implementação da Mobilidade 4.0 nos centros urbanos e seus impactos a partir da perspectiva humana. Para tal, foi feita uma revisão sistemática da bibliografia já publicada sobre este tema nos repositórios ERICH Plus, Google Scholar, DOAJ, Sci-Elo e publicações na internet em geral durante o período de 01 de outubro de 2019 a 30 de junho de 2023. Conclui-se, após análise e discussão de 56 publicações, que a Mobilidade 4.0 ainda é um cenário distante da realidade por limitações das tecnologias-base e postura dos seres humanos frente ao próprio sistema de trânsito, automação ou dilemas éticos e morais.

PALAVRAS-CHAVE: mobilidade 4.0; interação humano-máquina; inteligência artificial.

Structural integrity monitoring using artificial intelligence: Challenges, advances and applications

ABSTRACT: Over centuries of development, it has been possible for human beings to create and develop new tools that eliminate mobility needs within and outside the urban environment. Several technologies are being converged towards a new type of mobility: Mobility 4.0. This article aims to discuss the main obstacles to the implementation of Mobility 4.0 in urban centers and its impacts from a human perspective. Thus, a systematic review of the bibliography already published on this topic was carried out in the ERICH Plus, Google Scholar, DOAJ, Sci-Elo repositories, and publications on the internet in general during the period from October 1, 2019, to June 30, 2023. As a conclusion, after analysis and discussion of 56 publications, that Mobility 4.0 is still a scenario far from reality due to limitations of base technologies and the attitude of human beings towards the traffic system itself, automation, or ethical and moral dilemmas.

KEYWORDS: mobility 4.0; human-machine interaction; artificial intelligence.

* Universidade do Estado de Minas Gerais/ ** Universidade do Estado de Minas Gerais*** Universidade Federal de Juiz de Fora

Introdução

Ao longo de séculos de desenvolvimento tecnológico, foi possível ao ser humano criar e desenvolver novas ferramentas que pudessem suprir necessidades de mobilidade dentro e fora do ambiente urbano. Os automóveis, por exemplo, simbolizam o progresso ocidental e a liberdade de movimento que nasceram do desejo de se deslocar mais rápido e ir mais longe sem a dependência de cavalos (Redshaw, 2008; Pereira; Botelho, 2018).

O automóvel é um dos resultados permitidos graças à Segunda Revolução Industrial quando, uma vez consolidada a mecanização dos processos industriais, ocorreu a transição do modal energético do carvão e do vapor para a eletricidade e o petróleo (Hermann; Pentek; Otto, 2015; Perasso, 2016). Os automóveis, sejam eles de uso particular ou compartilhado, moldaram a forma como as cidades foram sendo projetadas e ampliadas. Muito do crescimento de determinadas nações só foi possível graças ao automóvel, bem como grande parte do desenvolvimento dos indivíduos está atrelado à existência e interação com tal objeto (Redshaw, 2008).

Se a força que impulsionou a Primeira Revolução Industrial foram os braços mecânicos impulsionados pelo vapor, a Quarta Revolução Industrial trará uma rede de sistemas digitais integrados, controlando fábricas inteiras com mão de obra robótica (Perasso, 2016). A Quarta Revolução Industrial nasce baseada na transformação digital ocorrida no início do século XXI e tem como pilar a Internet onipresente, sensores ultrassensíveis, Inteligência Artificial (IA) e aprendizagem de máquina (Schwab, 2016). A convergência de tecnologias à exemplo da automação, Internet das Coisas (IoT) e computação em nuvem, transformará uma planta fabril tradicional em uma “indústria inteligente” ou “Indústria 4.0” (Perasso, 2016; Miyasaka; Fabricio; Paoletti, 2019). Essas “indústrias inteligentes”, geridas e controladas de maneira autônoma e descentralizada, permitirão uma nova geração de valor, novos modelos de negócios focados em parâmetros globais e flexíveis de demanda (Schwab, 2016).

A convergência de automação avançada, IoT e computação em nuvem tem o poder de transformar uma indústria em uma “Indústria 4.0”, a aplicação dessas mesmas tec-

nologias na mobilidade urbana também serão capazes de criar uma “Mobilidade 4.o”. A Mobilidade 4.o, quando implementada, estará acessível a todos a qualquer momento e local. Sua disponibilidade, custos e manutenção estarão ligados às demandas personalizadas e individuais de seus usuários (Flugge, 2017). Deste modo, a Mobilidade 4.o será sustentada na completa automatização da maioria dos modais de transporte, na alta dependência de IA e equipamentos de alta tecnologia integrada aos veículos. Um exemplo desta mudança de paradigma é uma das linhas do metrô parisiense que, desde a última década do século XX, dispensa a necessidade de um operador humano (Augé, 2010).

No entanto, há um custo social relacionado ao atual modelo de mobilidade urbana dos grandes centros urbanos. Este custo social se traduz, por exemplo, em ferimentos e mortes devido a acidentes, congestionamentos, poluição ambiental e a imposição de rodízios de veículos (Mitrea; Kyamakya, 2017). Muitas cidades são despreparadas para acolher e gerenciar mais automóveis e diferentes modais de deslocamento, o que acaba por influenciar severamente o modo de vida da sociedade local (Silva, 2013). Com a robustez das tecnologias, que juntas viabilizam a Quarta Revolução Industrial, fica cada vez mais evidente que o futuro da mobilidade poderá ser feito por automóveis autônomos e, por consequência, os usuários desses modais tornar-se-ão apenas um elemento passivo no controle destes dentro da cadeia de transporte terrestre.

Diante de todo o contexto mencionado, esta revisão sistemática objetiva identificar e discutir os principais obstáculos para a implementação da Mobilidade 4.o nos centros urbanos e seus impactos a partir da perspectiva humana.

Materiais e métodos

Esta pesquisa utilizou o método indutivo e apoiou-se em uma revisão sistemática da literatura já publicada sobre o tema. Tal abordagem metodológica se justifica pelo fato da temática da Mobilidade 4.o ainda ser configurada como uma área de pesquisa em vias consolidação e ainda haver pouco material em língua portuguesa sobre o assunto. As palavras-chave utilizadas foram: interação humano-máquina, mobilidade inteligente, mobilidade urbana, veículos autônomos, transporte autônomo, cidade inteligente e quarta revolução industrial nos idiomas português e inglês.

Os critérios de inclusão usados na seleção da bibliografia foram: a) ter sido publicado nos últimos 25 anos; b) conter elementos que definiam os conceitos básicos das áreas de Inteligência Artificial, Mobilidade Urbana, Mobilidade Inteligente, Veículos Autônomos, Interação Humano-Máquina, Fatores Humanos e Quarta Revolução Industrial; c) apresentar problemas a respeito da automação de atividades humanas; d) conter elementos que registram aspectos sociais e individuais do ser humano moderno.

Já os critérios de exclusão da bibliografia utilizados foram: a) incluir conclusões consideradas ultrapassadas por publicações mais recentes; b) haver sobreposição de conteúdo ou ter conteúdos duplicados com outras publicações; c) não ser possível localizá-la por meio do uso de palavras-chave no idioma inglês ou português.

A busca ocorreu junto aos repositórios ERICH Plus, Google Scholar, DOAJ, Sci-Elo e publicações na internet em geral durante o período de 01 de outubro de 2019 a 30 de junho de 2023. O processo de levantamento e aceitação do material utilizado na revisão sistemática contou com três revisores que utilizaram um processo em cinco etapas:

i) busca feita online nos repositórios selecionados e, através da análise dos títulos, organização manual dos potenciais achados em uma planilha eletrônica; ii) leitura dos resumos e abstracts para verificar se atendiam os critérios de aceitação estabelecidos; iii) filtro com relação a data de publicação, além de se eliminar aqueles publicados anteriormente à data de corte estabelecida; iv) leitura dos artigos e das publicações mais relevantes em sua integralidade para verificar se atendiam aos critérios de inclusão; v) seleção das publicações que tivessem o aceite de dois dos três revisores para serem consideradas nesta revisão sistemática.

Resultados

da seleção das publicações

A busca considerou aptos para análise 29 artigos publicados em periódicos, 9 livros, 9 capítulos de livros, 4 artigos em revistas especializadas, além de artigos publicados em portais na internet (2), artigos publicados em jornais (1), normas técnicas (1) e dissertações de mestrado (1) (Figura 1).

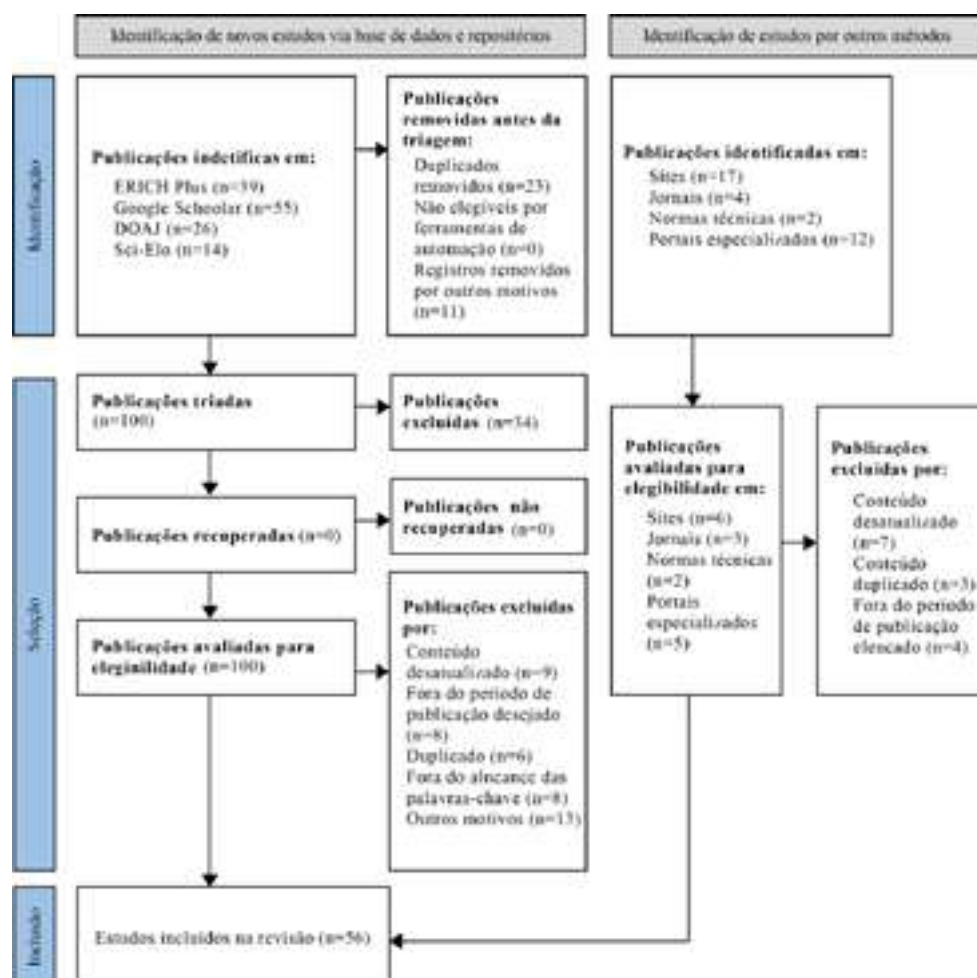


Figura 1 – Fluxograma PRISMA do processo de seleção de estudos e publicações.

Características das publicações

As 56 publicações encontradas foram organizadas de acordo com os macro grupos: a) Quarta Revolução Industrial; b) Fatores Humanos e Interações Humano-Máquina; c) Mobilidade Urbana. Os tópicos de agrupamento permitiram a organização dos assuntos como apresentado nos resultados (Quadro 1).

Material bibliográfico levantado para o estudo				
Autor (Ref.)	Ano	Título da Obra	Características	Macro grupo
Abuelsamid	2016	Tesla autopilot fatality shows why LIDAR and v2v will be necessary for autonomous cars.	Artigo em revista especializada	Quarta Revolução Industrial
Augé	2010	Por uma antropologia da mobilidade	Livro	Mobilidade Urbana
Azmat	2015	Impact of autonomous vehicles on urban mobility	Dissertação de mestrado	Mobilidade Urbana
Banyai Pereira et al.	2022	Factores humanos, inteligencia artificial y autos autónomos: Perspectivas para una implementación compleja	Artigo publicado em periódico científico	Fatores Humanos e Interações Humano-Máquina
Barbosa	2018	Trânsito como espaço social: personalidades e comportamentos	Artigo publicado em portal na internet	Mobilidade Urbana
Brooks	2014	O animal social: a história de como o sucesso acontece	Livro	Fatores Humanos e Interações Humano-Máquina
Campana et al.	2017	The Importance of Specific Usability Guidelines for Robot User Interfaces	Artigo publicado em periódico científico	Fatores Humanos e Interações Humano-Máquina
Davies et al.	2019	Are We There Yet? A Reality Check on Self-Driving Cars	Artigo em revista especializada	Fatores Humanos e Interações Humano-Máquina
Dogson	2016	Why Mercedes plans to let its self-driving cars kill pedestrians in dicey situations	Artigo em revista especializada	Fatores Humanos e Interações Humano-Máquina
Material bibliográfico levantado para o estudo (continuação)				
Autor (Ref.)	Ano	Título da Obra	Características	Macro grupo

Endsley	1996	Automation and situation awareness	Capítulo de livro	Fatores Humanos e Interações Humano-Máquina
Fagnant <i>et al.</i>	2013	Preparing a nation for autonomous vehicles: Opportunities, barriers and policy recommendations	Norma técnica	Mobilidade Urbana
Folcher <i>et al.</i>	2007	Homens, artefatos, atividades: perspectiva instrumental.	Capítulo de livro	Fatores Humanos e Interações Humano-Máquina
Floridi	2014	The 4th revollution: How the infosphere is reshaping human reality	Livro	Quarta Revolução Industrial
Flügge	2017	Introduction	Capítulo de livro	Quarta Revolução Industrial
Flügge <i>et al.</i>	2017	Impact of Smart Mobility on Existing Sectors	Capítulo de livro	Mobilidade Urbana
Hancock <i>et al.</i>	2013	Human-automation interaction research: past, present, and future	Artigo publicado em periódico científico	Fatores Humanos e Interações Humano-Máquina
Hermann <i>et al.</i>	2015	Design Principles for Industrie 4.0 Scenarios: A Literature Review	Artigo publicado em periódico científico	Quarta Revolução Industrial
Hoc	2007	Gestão da situação dinâmica	Capítulo de livro	Fatores Humanos e Interações Humano-Máquina
Hoffmann	2005	Comportamento do condutor e fenômenos psicológicos	Artigo publicado em periódico científico	Fatores Humanos e Interações Humano-Máquina
Hohenberger <i>et al.</i>	2016	How and why do men and women differ in their willingness to use automated cars? The influence of emotions across different age groups	Artigo publicado em periódico científico	Fatores Humanos e Interações Humano-Máquina
Hollnagel	2014	Human factors/ergonomics as a systems discipline? "The human use of human beings" revisited.	Artigo publicado em periódico científico	Fatores Humanos e Interações Humano-Máquina
Material bibliográfico levantado para o estudo (continuação)				

Autor (Ref.)	Ano	Título da Obra	Características	Macro grupo
Janssen <i>et al.</i>	2019	History and future of human-automation interaction.	Artigo publicado em periódico científico	Fatores Humanos e Interações Humano-Máquina
Jones	2017	Mobility – Change for the Good	Capítulo de livro	Mobilidade Urbana
Jordan <i>et al.</i>	1999	Computational intelligence.	Livro	Quarta Revolução Industrial
Kaber	2017	Issues in Human–Automation Interaction Modeling: presumptive aspects of frameworks of types and levels of automation	Artigo publicado em periódico científico	Fatores Humanos e Interações Humano-Máquina
Lapidus <i>et al.</i>	2023	Transformation of business models in the automotive industry in the context of the development of unmanned technologies	Artigo publicado em periódico científico	Quarta Revolução Industrial
Larica	2003	Design de automóveis: a arte em função da mobilidade	Livro	Mobilidade Urbana
Lee	2008	Review of a pivotal human factors article: “humans and automation: use, misuse, disuse, abuse”	Artigo publicado em periódico científico	Fatores Humanos e Interações Humano-Máquina
Magagnin	2008	A percepção do especialista sobre o tema mobilidade urbana.	Artigo publicado em periódico científico	Mobilidade Urbana
Marín-León <i>et al.</i>	2003	Comportamentos no trânsito: um estudo epidemiológico com estudantes universitários	Artigo publicado em periódico científico	Mobilidade Urbana
Maurer	2016	Introduction	Capítulo de livro	Quarta Revolução Industrial
Mcbride <i>et al.</i>	2019	Apple, Google, and Facebook Are Raiding Animal Research Labs	Artigo em revista especializada	Quarta Revolução Industrial
Medeiros	2018	Estresse e comportamentos de risco no trânsito	Artigo publicado em periódico científico	Mobilidade Urbana
Mitrea <i>et al.</i>	2017	(How) will autonomous driving influence the future shape of city logistics?	Artigo publicado em periódico científico	Mobilidade Urbana
Material bibliográfico levantado para o estudo (continuação)				
Autor (Ref.)	Ano	Título da Obra	Características	Macro grupo

Parikh <i>et al.</i>	2017	Autonomous cars using embedded systems and spherical wheels	Artigo publicado em periódico científico	Fatores Humanos e Interações Humano-Máquina
Norman	2008	Design emocional	Livro	Fatores Humanos e Interações Humano-Máquina
Ososky <i>et al.</i>	2014	Determinants of system transparency and its influence on trust in and reliance on unmanned robotic systems	Artigo publicado em periódico científico	Fatores Humanos e Interações Humano-Máquina
Parasuraman <i>et al.</i>	1997	Humans and Automation: use, misuse, disuse, abuse.	Artigo publicado em periódico científico	Fatores Humanos e Interações Humano-Máquina
Parikh <i>et al.</i>	2017	Autonomous cars using embedded systems and spherical wheels	Artigo publicado em periódico científico	Fatores Humanos e Interações Humano-Máquina
Perasso	2016	O que é a 4ª revolução industrial e como ela deve afetar nossas vidas.	Artigo publicado em Jornal	Quarta Revolução Industrial
Pereira <i>et al.</i>	2018	Design de Interação: fatores humanos e os carros autônomos.	Artigo publicado em periódico científico	Fatores Humanos e Interações Humano-Máquina
Piao <i>et al.</i>	2016	Public Views towards Implementation of Automated Vehicles in Urban Areas	Artigo publicado em periódico científico	Mobilidade Urbana
Prakken <i>et al.</i>	2015	Law and logic: A review from an argumentation perspective.	Artigo publicado em periódico científico	Mobilidade Urbana
Redshaw	2008	In the company of cars: driving as a social and cultural practice (Human factors in road and rail transport).	Livro	Fatores Humanos e Interações Humano-Máquina
Material bibliográfico levantado para o estudo (conclusão)				
Autor (Ref.)	Autor (Ref.)	Autor (Ref.)	Autor (Ref.)	Autor (Ref.)

Riley	1996	Operation reliance on automation: theory and data	Capítulo de livro	Fatores Humanos e Interações Humano-Máquina
Schank	1990	What is AI, anyway?	Capítulo de livro	Quarta Revolução Industrial
Schiller	2016	Automated and Connected Vehicles: High Tech Hope or Hype?	Artigo publicado em periódico científico	Quarta Revolução Industrial
Schwab	2018	Smart roads could protect us from self-driving cars crashes	Artigo publicado em portal na internet	Mobilidade Urbana
Schwab	2016	A quarta revolução industrial	Livro	Quarta Revolução Industrial
Silva	2013	Mobilidade urbana: os desafios do futuro	Artigo publicado em periódico científico	Mobilidade Urbana
Silveira <i>et al.</i>	2013	Transporte público, mobilidade e planejamento urbano: contradições essenciais	Artigo publicado em periódico científico	Mobilidade Urbana
Teixeira	1990	O que é inteligência artificial	Livro	Quarta Revolução Industrial
Teixeira	2011	Mente, cérebro e cognição	Livro	Quarta Revolução Industrial
Zhang <i>et al.</i>	2001	Heuristic search in artificial intelligence.	Artigo publicado em periódico científico	Quarta Revolução Industrial

Fonte: Elaborado pelos autores.

Sobre a mobilidade 4.0

A mobilidade urbana “pode ser definida como um atributo relacionado aos deslocamentos realizados por indivíduos nas suas atividades de estudo, trabalho, lazer e outras” (Magagnin; Silva, 2008, p.26). Segundo Azmat (2015), a “Mobilidade 4.0” é uma nova forma de deslocamento urbano, descrita como “mobilidade inteligente” na Quarta Revolução Industrial.

A Quarta Revolução é marcada por um desenvolvimento horizontal, a partir do qual as pessoas poderão se conectar entre si, com outros objetos e objetos poderão se conectar a outros objetos, transmitindo suas demandas e obtendo o devido retorno (Floridi, 2014). A forma como a estrutura urbana se configura guia o desenvolvimento dos meios de transporte (Larica, 2003). Uma vez que as montadoras de veículos têm incorporado, cada vez mais, itens eletrônicos de sensoriamento e Internet, espera-se que o mesmo ocorra com a infraestrutura das cidades. Dessa forma, estabelecida uma conexão sólida entre os dados emitidos/recebidos pelo ambiente e os dados emitidos/recebidos pelos veículos, a Mobilidade 4.0 estará em funcionamento.

Produtos da Quarta Revolução Industrial, a Internet onipresente, Computação em Nuvem e dispositivos móveis trabalharão em conjunto para que pessoas e empresas tomem melhores decisões de transporte baseado em custo, velocidade, saúde e meio-ambiente (Jones, 2017). Nesse sentido, a Mobilidade 4.0 tenderá a mudar o modelo de negócios da indústria automobilística que, ao invés de somente vender, essas empresas preocupar-se-ão em fornecer serviços de mobilidade (Lapidus; Shorokhova, 2023). Tais empresas tornar-se-ão provedores de mobilidade uma vez que os veículos poderão contar com sistemas avançados de IA para guiá-los, gerenciar frotas e tomar as melhores decisões para que os usuários sejam transportados do ponto A ao ponto B em segurança (Flugge; Pfriemer, 2017).

Sobre os fatores humanos e as novas tecnologias para a mobilidade 4.0

A área de Fatores Humanos é responsável pelo debate dos aspectos, físicos e psicológicos, que influenciam na execução de tarefas onde há manipulação de ferramentas, dispositivos e interações com sistemas e/ou máquinas (Larica, 2003). O estudo acerca de Fatores Humanos tem como objetivo encontrar soluções para problemas de ordem prática, obviamente, o impacto de novas tecnologias no dia-a-dia dos seres humanos acaba por demandar novas soluções para problemas até então inexistentes (Hollnagel, 2014).

Ao considerar as novas tecnologias trazidas pela Mobilidade 4.0, percebe-se a predominância de sistemas e veículos capazes de tomar decisões e comportarem-se no trânsito de forma a seguir regras previamente programadas sem, ou com limitada, intervenção humana (Maurer, 2016). Neste cenário, os seres humanos terão que interagir entre si e com sistemas automatizados ao mesmo tempo que precisarão lidar com as interações entre os próprios sistemas automatizados e destes sistemas com o ambiente (Floridi, 2014; Banyai Pereira; Dias Botelho, 2021).

Sistemas automatizados necessitam ser concebidos observando critérios relacionados à interação humano-máquina, nos quais o usuário comunica-se com a máquina por meio de uma interface (Folcher; Rabardel, 2007). Dentro de um veículo moderno, por exemplo, a interface ocorre por meio de softwares, assim como por diversos controles e comandos que controlam funções e ficam dispostos fisicamente em um painel (Larica, 2003).

A Mobilidade 4.0 trará uma gama de agentes automatizados que atuarão como intermediários entre o desejo de locomoção das pessoas e seus destinos finais. Assim como ocorre em interações diretas entre humanos e sistemas automatizados, aqui também faz-se essencial o uso de interfaces eficientes que comuniquem de forma natural o andamento da atividade para o usufruto da automação (Campana; Quaresma, 2017). Por se tratarem de atividades dinâmicas e complexas, nas quais são predominantes a presença de diversas variáveis cujos agentes envolvidos não possuem total controle da ação, haverá riscos e incertezas (Hoc, 2007). Consequentemente, cabe ao sistema a tarefa de transparecer/informar/mostrar/atualizar o estado da execução da atividade a todo o momento.

Análise e discussão

Diversos desafios se relacionam com a questão da mobilidade autônoma terrestre. O desenvolvimento tecnológico que possibilita esse tipo de mobilidade terá impacto significativo na sociedade humana e na forma como ela se organiza no trânsito. Isto posto, nota-se um relevante desafio para a aceitação da mobilidade autônoma: os fatores humanos sob a ótica de como a IA impactará na organização social vigente. Isso se deve pelo fato de o trânsito representar o movimento de múltiplos agentes e interesses que, por vezes, podem entrar em conflito (Berwing, 2013).

Um ponto crucial em relação ao desafio relevante mencionado estaria em encarar o advento dos veículos autônomos de uma forma diferente, pois não se trataria mais de “quando” a mobilidade autônoma virá, mas “de que forma” e “para quem” ela surgirá primeiro (Davies; Marshall, 2019). Observa-se, novamente, que os veículos autônomos são componentes dentro de um macro sistema de mobilidade inteligente e conectada. Uma mobilidade autônoma altamente dependente de Internet e de circuitos eletrônicos levanta preocupações sobre a segurança de todo o sistema de trânsito (Fagnant; Kockelman, 2013).

Assim, ao entender a complexidade que envolve a mobilidade autônoma, o trânsito e os diversos agentes envolvidos, explora-se, neste artigo, uma abordagem que contempla os obstáculos principais para a mobilidade 4.0 por meio de dois fatores macro: externos e internos ao ser humano. Tais fatores são complementares entre si e, em paralelo, refletem, o maior desafio para a implementação da mobilidade autônoma: o ponto de vista do usuário.

5.1 FATORES DE INFLUÊNCIA EXTERNOS AO SER HUMANO

O primeiro fator de influência externo ao ser humano levado em consideração trata dos problemas envolvendo o trânsito dentro e fora dos centros urbanos. O trânsito sofre influência de elementos locais, então nota-se a importância dos fatores externos ao ser humano na forma como estes se apresentam, sejam eles culturais, estruturais, sociais ou econômicos (Silveira; Cocco, 2013; Barbosa, 2018). É inegável a influência do ambiente e da sociedade em relação ao indivíduo e suas atividades. Independentemente do modal de transporte, é necessário que haja uma infraestrutura condizente e, quanto mais complexo for o meio de transporte, mais equipamento de suporte, manutenção e segurança o sistema demandará (Larica, 2013).

O segundo fator de influência externo ao ser humano está relacionado com a aleatoriedade do ambiente de uso. O trânsito possui uma dinâmica própria, está envolto em simbolismos e comunicações específicas. O mundo da mobilidade urbana é complexo em suas diversidades e por isso imprevisível. Robôs autônomos ainda possuem custos elevados e demasiadamente sensíveis para as demandas exigidas a um motorista humano (Davies; Marshall, 2019). Um veículo autônomo talvez possa, em situações simples, optar por decisões assertivas. No entanto, a garantia de uma segurança total ainda se mostra como sendo algo fora de alcance (Shiller, 2016).

É comum que sejam ressaltados os benefícios de veículos autônomos rodando nas ruas, seja para apontar possíveis reduções nos acidentes, seja para apontar um emprego eficaz de energia para a propulsão (Pereira; Botelho, 2018). No entanto, a própria coexistência entre o conceito atual de mobilidade com o conceito de mobilidade autônoma não é abordada na literatura disponível sobre o assunto. Um dos primeiros problemas acontece na incapacidade dos objetos que compõem o cenário urbano de se comunicarem, informarem suas ações e estados.

Tal comunicação precisa ser estabelecida para que haja uma efetiva tomada de decisões dos veículos autônomos no momento em que suas câmeras e radares lhes fornecerem informações conflitantes ou insuficientes (Abuelsamid, 2016). Mesmo dotados de sensores, os veículos característicos da mobilidade autônoma teriam dificuldade para distinguir entre situações similares como, por exemplo, buracos ou poças de água suja, sombras de estruturas e diferenças na tonalidade da manta asfáltica (Shiller, 2016).

Se a presença de objetos estáticos no ambiente já parece colocar em xeque a implementação de veículos autônomos terrestres, quando se adicionam elementos dinâmicos como, por exemplo, outros veículos e pedestres, o cenário para essa possível implementação parece ainda mais desanimador. Como ressalta Schiller (2016), uma vez em circulação, esses veículos interagirão com objetos não gerenciados por IA e seus algoritmos poderão ser ineficazes em captar, analisar e processar as possibilidades de interações com esses objetos na velocidade necessária. Tendo em vista a tendência desses veículos autônomos em terem como prioridade a segurança e conforto de seus passageiros, caberá à infraestrutura das cidades gerenciarem de forma segura seus pedestres, animais e ciclistas (Schwab, 2018). Sob essa perspectiva, a aleatoriedade do ambiente urbano, no mínimo, precisará ser monitorada para auxiliar os veículos autônomos no planejamento de suas ações.

Por mais avançados que pareçam, os sensores óticos dos veículos autônomos ainda são equiparados à visão humana e estão sujeitos a interpretações incorretas caso a captação da luz seja precária (Abuelsamid, 2016). Existe ainda uma limitação nos sistemas de GPS (sistema de posicionamento global) e nos sistemas que mapeiam o entorno do veículo, impedindo-os de se atualizarem em tempo real sobre as dinâmicas humanas e do clima sobre a infraestrutura da pista (Schiller, 2016). Uma vez que a comunicação entre veículos autônomos e entre esses e o ambiente se dará via internet, caberá a todos (veículos e ambiente) comunicar a pedestres e ciclistas o estado dessa dinâmica (Schwab, 2018).

Observa-se que ainda é cedo para confiar aos veículos sem motorista todas as demandas que a mobilidade autônoma exige. A infraestrutura da pista precisa estar tão tecnologicamente atualizada quanto os veículos que trafegam por ela. Como sugerem Jordan e Russel (1999), sistemas autônomos perdem consideravelmente parte de suas capacidades ao lidarem com ambientes dinâmicos com variáveis em excesso. Para que elementos cujos comportamentos são aleatórios, como por exemplo, ciclistas, pedestres, animais e condições climáticas; é preciso multiplicar os agentes de coleta, análise e processamento de dados em toda a cadeia de transporte.

O terceiro fator de influência externo ao ser humano está relacionado com a IA que é o principal pilar da mobilidade autônoma e a complexidade dessa área inicia em sua própria definição. Esta vai para além das áreas de engenharia de softwares ou matemática, abarcando áreas como linguística, psicologia, filosofia, etc. (SCHANK, 1990). No entanto, ainda a associamos a um ramo da ciência capaz de reproduzir as capacidades cognitivas e planejar reações tidas, até então, como exclusivas da espécie humana (Teixeira, 1990).

É sabido que a reprodução da inteligência humana por uma máquina ainda é impossível de ser atingida, dessa maneira há um esforço para o desenvolvimento e aperfeiçoamento de máquinas capazes de aprenderem a executar sozinhas determinadas demandas (Mcbride; Vance, 2019). Isso se mostra desafiador uma vez que um dos objetivos da pesquisa em IA é descobrir a natureza da própria inteligência (Schank, 1990). Para se fazer uma máquina inteligente é preciso entender o que é inteligência e de onde ela se origina. Sendo assim, faz-se importante ressaltar a notável ausência de habilidade das máquinas na propensão para encarar fatos e ideias de um

ponto de vista puramente racional. Essa ausência de racionalidade pura ainda é um abismo significativo entre máquinas autônomas e os seres humanos (Teixeira, 2011).

De acordo com Augé (2010, p.10), “a ciência não para de progredir em um ritmo acelerado (somos incapazes de dizer qual será o estado de nossos conhecimentos em trinta anos)”. Em determinado momento existirá uma realidade na qual máquinas autônomas serão comuns e se encarregarão de todas as atividades perigosas e tediosas (Norman, 2008). Enquanto não se obtiver uma infraestrutura dotada de IA capaz de auxiliar um veículo também possuidor de IA, não será possível confiar toda a resolução de problemas que o trânsito exige a um só agente inteligente.

Um agente inteligente focado na resolução de problemas gerais tende a ser menos eficiente se comparado a um outro designado para a resolução de problemas específicos (Zhang; Dechter; Korf, 2001). Enquanto não for possível uma rede que conecte diferentes agentes inteligentes especializados, os usuários serão elementos-chave na forma como tais máquinas lidarão com novas situações (Pereira ; Botelho, 2018).

Espera-se que uma IA no controle de um veículo autônomo, ao imitar como seres humanos imprimem direção e sentido, seja capaz de evitar colisões e seguir um determinado caminho. Por isso é aguardado que essa IA também seja capaz de imitar todas as atividades mentais ligadas à condução veicular (Teixeira, 1990). Infelizmente, mesmo com todo esforço obtido em 60 anos de pesquisas e desenvolvimento, os computadores com maior capacidade de processamento de dados são capazes de emular uma fração do poder de análise do cérebro de um camundongo médio (Mcbride; Vance, 2019).

Os avanços graduais nas áreas que trabalham com IA fomentaram a criação de realidades otimistas quanto à capacidade dos computadores de se igualarem à inteligência humana. Apesar das críticas a esse otimismo, tais avanços apontam para um desafio a ser superado para que se possa ter algo semelhante à mente humana (Teixeira, 1990).

Fatores de influência internos ao ser humano

Assim como os fatores de influência externas ao ser humano, existem aqueles de influência interna. O primeiro deles versa sobre o comportamento, emoções e cognição humana. A cognição humana é capaz de dar ao indivíduo uma visão ampla do

mundo à sua volta, no entanto, são as emoções que o fazem agir de maneira repentina antes que seja possível avaliar todo o contexto (Norman, 2008). Segundo Barbosa (2018), o comportamento de um indivíduo em conjunto com as emoções impacta na forma como esse indivíduo age enquanto dirige. A IA responsável pelas tomadas de decisões de um agente autônomo deverá respeitar as leis de trânsito local, ao passo que deverá lidar com agentes humanos, que não necessariamente seguem de maneira correta as mesmas leis (Hoffmann, 2005).

Veículos autônomos precisarão de algoritmos complexos para prever o comportamento de um carro guiado por um ser humano ao passo que avalia em que medida é possível infringir uma lei de trânsito para evitar acidentes. Existe uma relação entre uma maior ocorrência de acidentes entre pessoas que têm pouco apreço pelas leis de trânsito se comparada com aquelas que veem benefícios nelas (Marín-León; Vizzotto, 2003). Há ainda a possibilidade de os veículos autônomos aprenderem sozinhos a partir do reconhecimento dos padrões de condução dos motoristas humanos. Os diversos agentes humanos presentes no trânsito são complexos por natureza, além do conhecimento das leis de trânsito e sua disposição em cumpri-las, seu estado emocional influencia severamente no seu comportamento (Barbosa, 2018).

Homens e mulheres parecem encarar a possibilidade de se locomover por meio de veículos autônomos de maneiras diferentes. Mulheres tendem a elencar emoções negativas como, por exemplo, a ansiedade, ao pensar em veículos autônomos; ao passo que homens têm a tendência de elencar emoções positivas como, por exemplo, o prazer (Hohenberger; Sporrle; Welp, 2016). Apesar da abertura que a maioria das pessoas demonstra ter sobre veículos autônomos, as iniciativas abertas ao grande público costumam operar em baixas velocidades e com pouca interação com o trânsito típico dos grandes centros urbanos. Isso quebra as expectativas das pessoas quanto à capacidade dessa tecnologia em gerar resultados e ajudar na redução do consumo de combustível (Piao et al., 2016).

Para Norman (2008), um dos alicerces que fundamentam a base das relações humanas é a confiança. No caso de veículos autônomos que venham frustrar as expectativas de seus usuários, tal relação consistirá em um ataque à credibilidade, pois tais veículos não desempenham suas funções de acordo com o que se espera deles. Isso se explica pelo fato da mente humana criar modelos e padrões de como as coisas funcio-

nam. Portanto, quando há diferenças entre o modelo mental gerado e a realidade, as pessoas tendem a demonstrar emoções negativas (Brooks, 2014).

Se um veículo autônomo seguir rigidamente as leis de trânsito, poderá apresentar desempenho inferior graças à postura de motoristas humanos à sua volta. Motoristas humanos poderão, por exemplo, ignorar a preferência de passagem de um veículo autônomo fazendo com que a IA do veículo evite forçar um avanço, o que aumentaria as chances de colisão. O ambiente do qual participam carros autônomos e seres humanos é caótico, tendendo a levar seus usuários a situações estressantes gerando um sistema retro alimentador, o qual, ao estressar o componente humano, influencia o seu comportamento levando-o a gerar mais desordem e mais estresse (Medeiros et al., 2018).

Arelada ao comportamento e às emoções está a cognição humana, que desempenha papel relevante na interação das pessoas com esse novo conceito de mobilidade. Segundo Larica (2003), para que se possa lidar com aparelhos tecnológicos contemporâneos é preciso que se desenvolva uma série de linguagens e de repertório diversificados. Isso graças às múltiplas possibilidades que tais dispositivos oferecem, visto que nem sempre, a interação com esses aparelhos baseia-se na forma como se interagiria com modelos anteriores.

Dirigir não exige do motorista o conhecimento específico de como cada subsistema de propulsão trabalha em conjunto com outros para que o veículo se movimente (Ososky et al., 2014). No entanto, por mais simples que dirigir um automóvel possa parecer, a atividade requer certo treino e repertório, sobretudo quando se considera que situações adversas acontecem mesmo para aqueles profissionais bem treinados e em plena posse de suas faculdades mentais e físicas.

Percebe-se, então, uma carência de fronteiras que estabeleçam corretamente as interações de aspectos puramente humanos e artificiais (Mcbride; Vance, 2019). Veículos com tecnologia autônoma embarcada podem ser vistos como objetos altamente complexos. As ações e decisões desses veículos extrapolam a imagem mental criada a partir dos carros sem automação. Assim sendo, um veículo autônomo pode apresentar erros de diversas formas a ponto de frustrar as expectativas de seus ocupantes (Norman, 2008).

O segundo fator de influência interno diz respeito a interação humano-automatização. O desenvolvimento tecnológico, observado nos últimos anos, permitiu à automatização

expandir sua presença para além das fábricas. Para Floridi (2014), é notório o hábito humano em delegar, cada vez mais, à automação suas tarefas rotineiras, gestão de dados e tomadas de decisões.

É notável que mesmo havendo um sistema automatizado que garanta uma mobilidade autônoma para a maioria dos elementos presentes em um sistema viário, operar um veículo em vias públicas pode se mostrar desafiador. Existem muitas interações com objetos de comportamentos pouco previsíveis (outros veículos conduzidos exclusivamente por humanos, pedestres, animais, etc.) (Parikh; Joshi, 2017). Essas questões são de grande importância, haja vista a quantidade de usuários a bordo de veículos totalmente (ou parcialmente) automatizados que não possuem treinamento de como lidar com automação. O contato esporádico com esses veículos ou a falta de treinamento adequado pode alimentar expectativas incorretas frente às capacidades e limitações da automação ali presente (Janssen et al., 2019).

Segundo Kaber (2017), existe uma dificuldade no campo da interação humano-automação, relacionada ao desenvolvimento de modelos que prevejam o desempenho tanto do humano quanto do sistema. Diversos fenômenos da interação humano-automação observados em contextos industriais podem ser aplicados no caso da mobilidade autônoma tais como excesso de confiança no sistema automatizado, subutilização de funções e de recursos, assim como, negligência das limitações do próprio sistema.

A decisão de um agente humano por usar um sistema automatizado baseia-se na confiabilidade e precisão desse sistema (Parasuraman; Riley, 1997). Havendo uma percepção de que o sistema é confiável e preciso, o agente humano pode desenvolver uma complacência frente à automação. A complacência se caracteriza pelo excesso de confiança de um operador humano nas capacidades de um sistema automatizado em executar uma atividade, acarretando a perda da consciência na situação de atividade (Hancock et al., 2013). Quando o sistema falha, por qualquer motivo, o agente humano sem ciência da situação necessitará de mais tempo e esforço cognitivo para entender o status do sistema e o ambiente à sua volta para formular uma (re)ação corretiva adequada (Endsley, 1996).

Outro fenômeno típico da interação humano-automação é a subutilização de recursos e funções. Esse fenômeno é fruto das percepções do agente humano frente à confiabilidade e precisão de um sistema automatizado. No entanto, tende a ocorrer em

situações de alto risco nas quais as escolhas do agente humano geram consequências significativas (Riley, 1996). São exemplos da subutilização de recursos e funções ignorar avisos de segurança ou desativar notificações e alertas (Parasuraman; Riley, 1997). Esse comportamento de dúvida do agente humano em relação às capacidades do sistema automatizado pode sobrecarregar o operador e, a depender do grau de descrença na automação, induzi-lo ao erro.

A negligência das limitações do próprio sistema automatizado é outro fenômeno relevante nas interações humano-automação. Segundo Parasuraman e Riley (1997), este fenômeno está relacionado à tomada de decisão de um projetista ou gerente, pelo uso da automação em situações não apropriadas para tal. Essa negligência pode aparecer na falta de consideração do agente humano como parte da atividade executada e a sua intervenção não é considerada em caso de falhas do sistema (Hancock et al., 2013). Sistemas automatizados projetados para substituir agentes humanos podem ter seu desempenho comprometido por excluir intervenções de operadores em situações em que o sistema não é capaz de contornar problemas por conta própria (Lee, 2008).

Percebe-se, também, outro fenômeno que emerge nas interações entre humanos e sistemas automatizados: o abuso. Este se refere ao fenômeno do humano impor uma condição de trabalho para além das capacidades de um sistema automatizado. No contexto da mobilidade automatizada, o abuso pode ocorrer quando algum pedestre ou ciclista, por livre e espontânea vontade, executa uma manobra de risco sabendo que os veículos automatizados são programados para evitar acidentes, por exemplo. Ou quando esses mesmos pedestres ou ciclistas se colocam em situações de risco como forma de saciar uma curiosidade própria em saber quais estratégias serão tomadas pelo veículo automatizado para evitar um acidente.

Percebe-se, então, que os diferentes fenômenos ligados à interação humano-automação observados em setores industriais e militares encontram representações equivalentes em um cenário de mobilidade autônoma. Usuários poderão apresentar posturas diversas frente à alta automação dos veículos. Alguns podem demonstrar total complacência, total descrença e outros poderão se sentir instigados a testar os limites da automação. Além disso, fabricantes desses veículos podem desenvolver sistemas que eliminam o usuário das rotinas de execução da tarefa, dificultando a intervenção do usuário em caso de anomalias no desempenho seguro do veículo.

Por fim, o terceiro fator de influência interno está relacionado a questões éticas e morais. A sociedade contemporânea é dotada de dogmas e signos que podem representar um grande obstáculo para a implementação de uma mobilidade totalmente autônoma. Isto se dá pelo fato de existir um cenário de transição, no qual veículos autônomos interagirão com veículos não-autônomos em ambientes que podem conter, ou não, meios para emitir informações de auxílio à veículos e motoristas.

Esse cenário de transição não possui a maior parte das suas variáveis controladas ou previsíveis e, portanto, veículos autônomos necessitarão de capacidades de tomadas de decisões tão complexas quanto as capacidades humanas. Por consequência, a escolha por determinada atitude acaba por esbarrar em questões éticas e morais vigentes na sociedade na qual tal veículo estiver inserido.

Um primeiro indício seria a questão de como novos modelos de mobilidade vêm reforçando ainda mais os aspectos individualistas e pouco conectados com as características sociais e regionais do usuário (Augé, 2010). Cedo ou tarde, veículos autônomos estarão diante de dilemas morais como, por exemplo, atropelar pedestres para proteger a si mesmos e seus ocupantes (Bonnefon; Shariff; Rahwan, 2016). Os algoritmos que guiarão os veículos autônomos estão no limite técnico e ético acerca de qual é a melhor saída em caso de acidentes e falhas repentinas ou, ainda, de quais são as vidas que devem ser priorizadas (Schiller, 2016).

Veículos autônomos chegarão ao mercado com elevados preços, sendo adquiridos por uma parcela da população economicamente elevada. Tais compradores, dado o investimento expressivo nesses veículos, não se sentirão atraídos por veículos cujo computador central opte por sacrificá-lo em prol de um número qualquer de pedestres. Essa percepção parece encontrar respaldo na gerência do departamento de sistemas de auxílio ao motorista da Mercedes-Benz, que priorizará a segurança dos ocupantes do veículo em detrimento de outras (Dogson, 2016). Mundialmente, os sistemas jurídicos ainda carecem de precedentes sólidos para lidar com tais situações. Se parece difícil aplicar sentenças a seres humanos quando tomam decisões em situações como essas, pode-se considerar próximo do impossível aplicar sentenças quando o agente dessas decisões é um veículo inteligente.

A lei e a moral são um dos pilares para o bem-estar de uma sociedade. Na lei e na moral encontram-se recursos para garantir justiça e coesão social, uma vez que sua criação e aplicação carece de fundamentos lógicos e passíveis de justificativas (Prakken;

Sartor, 2015). Veículos que não necessitam de motoristas, assim como tantas outras máquinas automatizadas, estão se encarregando de tarefas complexas e cabe aos seus inventores programarem o modo como será a interação destes com o ambiente e as pessoas (Norman, 2008).

Até o presente momento, parece não haver restrições jurídicas que resguardecam a integridade humana nas interações do veículo autônomo com o ambiente e as pessoas. Isto deixaria os não-ocupantes de tais veículos à mercê da bússola moral de quem fabrica e programa tais máquinas. Em adição, as pessoas parecem ser contra a imposição de leis que obriguem as montadoras a produzir veículos tomadores de decisões utilitaristas, ao passo que reconhecem que leis muito restritivas podem atrasar a chegada de carros autônomos, o que contribuiria para a manutenção dos índices de acidentes fatais (Bonnefon; Shariff; Rahwan, 2016).

Considerações finais

Este artigo objetivou analisar os principais obstáculos para a implementação da Mobilidade 4.0 ao passo que se discutia seus impactos a partir da perspectiva humana. O recorte do trabalho focou na mobilidade urbana terrestre, sobretudo, nos veículos autônomos que tenderão a compartilhar vias públicas com pedestres, ciclistas, animais e outros motoristas. Através de uma revisão sistemática da literatura levantada, percebeu-se que a mudança de paradigma trazida pela Mobilidade 4.0 adiciona ao ambiente novas camadas de complexidade ao contexto urbano. Dessa maneira, um cenário de completa automatização da mobilidade urbana ainda se mostra distante, apesar dos esforços de diversos setores da indústria apontarem neste sentido.

Após análise, concluiu-se que o cenário de transição para uma mobilidade não-autônoma para uma Mobilidade 4.0 possui diversos desafios. Tais desafios estão relacionados à fatores externos ao ser humano e internos à ele. Os fatores externos incluem a alta aleatoriedade do ambiente de uso de veículos autônomos e as limitações que a IA tem ao lidar com múltiplas variáveis ao mesmo tempo. A solução para tais desafios parece caminhar no destino de áreas e vias exclusivas para veículos autônomos, onde haja alto controle sobre as variáveis que ali incidem enquanto são desenvolvidas IAs generalistas.

Já os fatores internos incluem os comportamentos, as emoções e a cognição humana atuantes na forma como seres humanos interagem com o trânsito e com sistemas automatizados. Questões éticas e morais também desempenham um papel importante na maneira como a sociedade lidará com falhas no sistema e veículos autônomos. A solução para estes desafios parece caminhar na reforma educacional para o trânsito, incluindo procedimentos de como interagir e operar veículos autônomos. Em adição,

Referências

- ABUELSAMID, Sam. *Tesla autopilot fatality shows why lidar and v2v will be necessary for autonomous cars*. Revista Forbes. Jul. 2016. Disponível em: <<https://www.forbes.com/sites/samabuelsamid/2016/07/01/first-tesla-autopilot-fatality-demonstrates-why-lidar-and-v2v-probably-will-be-necessary/#56cd5b892d91>>. Acesso em: 27 out. 2019.
- AUGÉ, Marc. *Por uma antropologia da mobilidade*. Limeiras: EDUFAL, 2010. 109p.
- AZMAT, Muhammad. *Impact of autonomous vehicles on urban mobility*. Dissertação (Mestrado) - Curso de Mestrado em Supply Chain Management, Institut Für Transportwirtschaft Und Logistik, Vienna University Of Economics And Business, Vienna, 2015. Disponível em: https://www.researchgate.net/publication/281243480_IMPACT_OF_AUTONOMOUS_VEHICLES_ON_URBAN_MOBILITY. Acesso em: 29 mar. 2020.
- BANYAI PEREIRA, S.; DIAS BOTELHO, R. *Factores humanos, inteligencia artificial y autos autónomos: Perspectivas para una implementación compleja*. Revista de Ciencia y Tecnología, v. 36, n. 1, p. 61–69, 2021. DOI: 10.36995/j.recyt.2021.36.008.
- BARBOSA, Lucicleide L. *Trânsito como espaço social: personalidades e comportamentos*. Psicologia.pt – O Portal dos Psicólogos. p. 1-9, 2018. Disponível em: <<https://www.psicologia.pt/artigos/textos/A1185.pdf>>. Acesso em 9 dez. 2019
- BROOKS, David. *O animal social: a história de como o sucesso acontece*. Rio de Janeiro: Objetiva, 2014. 496p.
- CAMPANA, Julia R.; QUARESMA, Manuela. *The Importance of Specific Usability Guidelines for Robot User Interfaces*. In: MARCUS, Aaron; WANG, Wentao (ed.). *Design, User Experience, and Usability: designing pleasurable experiences*. Designing Pleasurable Experiences. v.10289. (Lecture Notes in Computer Science - Part II). Proceedings of the 6th International Conference on Design, User Experience, and Usability, DUXU 2017. Vancouver: Springer, 2017. p. 471-483.
- DAVIES, Alex; MARSHALL, Aarian. *Are We There Yet? A Reality Check on Self-Driving Cars*. Revista Wired. Abr. 2019. Disponível em: <<https://www.wired.com/story/future-of-transportation-self-driving-cars-reality-check/>>. Acesso em: 14 out. 2019.
- DOGSON, Lindsay. *Why Mercedes plans to let its self-driving cars kill pedestrians in dicey situations*. Business Insider. Out. 2016. Disponível em
- ENDSLEY, Mica R. *Automation and situation awareness*. In: PARASURMAN, R.; MOULOUA, M., (Ed.). *Automation and human performance: theory and applications*. Mahwah: Lawrence Erlbaum, 1996. p.163-181.
- FAGNANT, Daniel, J.; KOCKELMAN, Kara, M. *Preparing a nation for autonomous vehicles: Opportunities, barriers and policy recommendations*. ENO Center for Transportation. Washington – DC. 2013. Disponível em: <<https://www.enotrans.org/etl-material/preparing-a-nation-for-autonomous-vehicles-opportunities-barriers-and-policy-recommendations/>>. Acesso em: 15 out. 2018.

FOLCHER, Vivian; RABARDEL, Pierre. Homens, artefatos, atividades: perspectiva instrumental. In: FALZON, P. *Ergonomia*. São Paulo: Blucher, 2007, p. 207-222.

FLORIDI, Luciano. *The 4th revollution: How the infosphere is reshaping human reality*. Nova Iorque: Oxford Press, 2014. 264p.

FLÜGGE, Barbara. Introduction. In: FLÜGGE, B. (ed.). *Smart Mobility – Connecting Everyone*, *Gewerbest: Springer Fachmedien Wiesbaden*. 2017, cap. 1, p. 1-3. http://dx.doi.org/10.1007/978-3-658-15622-0_1

FLÜGGE, Barbara; PFRIEMER, Heinrich. Impact of Smart Mobility on Existing Sectors. *Smart Mobility – Connecting Everyone*, p. 85, 2017. *Springer Fachmedien Wiesbaden*.

HANCOCK, Peter A. et al. Human-automation interaction research: past, present, and future. *Ergonomics in Design: The Quarterly of Human Factors Applications*, v. 21, n. 2, p. 9–14, abr. 2013. <https://doi.org/10.1177/1064804613477099>.

HERMANN, Mario; PENTEK, Tobias; OTTO, Boris. Design Principles for Industrie 4.0 Scenarios: A Literature Review. Working paper, p. 1-16, 2015. <https://doi.org/10.1109/HICSS.2016.488>

HOC, Jean-Michel. Gestão da situação dinâmica. In: FALZON, P. *Ergonomia*. São Paulo: Blucher, 2007, p. 443-454.

HOFFMANN, Maria H. Comportamento do condutor e fenômenos psicológicos. *Psicologia: Pesquisa & Trânsito*, v. 1, n. 1, p. 17-24, jul./dez. 2005.

HOHENBERGER, Christoph; SPÖRRLE, Matthias; WELPE, Isabell M.. How and why do men and women differ in their willingness to use automated cars? The influence of emotions across different age groups. *Transportation Research Part A: Policy and Practice*, v. 94, p. 374-385, dez. 2016. Elsevier BV. <http://dx.doi.org/10.1016/j.tra.2016.09.022>.

HOLLNAGEL, Erik. Human factors/ergonomics as a systems discipline? “The human use of human beings” revisited. *Applied Ergonomics*, v. 45, n. 1, p. 40-44, jan. 2014. Elsevier BV. <http://dx.doi.org/10.1016/j.apergo.2013.03.024>.

JANSSEN, Christian P. et al. History and future of human-automation interaction. *International Journal Of Human-Computer Studies*, v. 131, p. 99-107, nov. 2019. Elsevier BV. <http://dx.doi.org/10.1016/j.ijhcs.2019.05.006>.

JONES, Tim. Mobility – Change for the Good. In: FLÜGGE, B. (edit). *Smart Mobility – Connecting Everyone*. Wiesbaden: Springer Vieweg, 2017, p. 9-14. https://doi.org/10.1007/978-3-658-15622-0_2.

JORDAN, Michael I.; RUSSELL, Stuart. Computational intelligence. *The MIT encyclopedia of the cognitive sciences*. Cambridge: The MIT Press. 1999, p. 74-91

KABER, David B.. Issues in Human–Automation Interaction Modeling: presumptive aspects of frameworks of types and levels of automation. *Journal Of Cognitive Engineering And Decision Making*, v. 12, n. 1, p. 7-24, 19 out. 2017. SAGE Publications. <http://dx.doi.org/10.1177/1555343417737203>.

LAPIDUS, L. V.; SHOROKHOVA, V. N.. Transformation of business models in the automotive industry in the context of the development of unmanned technologies. *Intellect. Innovations. Investments*, n. 2, p. 19-33, 2023. Orenburg State University. <http://dx.doi.org/10.25198/2077-7175-2023-2-19>.

LARICA, Neville J. Design de automóveis: a arte em função da mobilidade. 1.ed. Rio de Janeiro: 2AB/

PUC-Rio, 2003. 216p.

LEE, John D. Review of a pivotal human factors article: "humans and automation: use, misuse, disuse, abuse". *Human Factors*, v. 50, n. 3, p. 404-410, 1 jun. 2008, Human Factors and Ergonomics Society. <https://doi.org/10.1518/001872008X288547>.

MAGAGNIN, Renata C.; SILVA, Antônio N. R. A percepção do especialista sobre o tema mobilidade urbana. *Transportes*, v. 16, n. 1, p. 25-35, 17 dez. 2008. *Lepidus Tecnologia*. <http://dx.doi.org/10.14295/transportes.v16i1.13>.

MARÍN-LEÓN, Leticia; VIZZOTTO, Marília Martins. Comportamentos no trânsito: um estudo epidemiológico com estudantes universitários. *Cad. Saúde Pública*, Rio de Janeiro, v. 19, n. 2, p. 515-523, Apr. 2003. <http://dx.doi.org/10.1590/S0102-311X2003000200018>.

MAURER, M. Introduction. In: MAURER, M.; GERDES, J. C.; Lenz, B.; WINNER, H. (Orgs). *Autonomous Driving: technical, legal and social aspects*. Berlin: Springer, 2016. cap. 1, p. 1-7. http://dx.doi.org/10.1007/978-3-662-48847-8_1.

MCBRIDE, Sarah; VANCE, Ashlee. Apple, Google, and Facebook Are Raiding Animal Research Labs. *Bloomberg*. Seção Hyperdrive. Jun. 2019. Disponível em: <<https://www.bloomberg.com/news/features/2019-06-18/apple-google-and-facebook-are-raiding-animal-research-labs?srnd=hyperdrive>>. Acesso em: 21 out. 2019.

MEDEIROS, E. L., et al. Estresse e comportamentos de risco no trânsito. *Temas em Saúde*, João Pessoa, v. 18 n. 1, p. 31-50, 2018. Disponível em: <<http://temasemsaude.com/wp-content/uploads/2018/04/18103.pdf>>. Acesso em: 09 dez. 2019.

MITREA, Oana; KYAMAKYA, Kyandoghere. (How) will autonomous driving influence the future shape of city logistics? *Istrazivanja I Projektovanja Za Privredu*, v. 15, n. 1, p. 45-52, 2017. Centre for Evaluation in Education and Science (CEON/CEES). <http://dx.doi.org/10.5937/jaes15-12178>.

MIYASAKA, E. L. ; FABRICIO, M. M. ; PAOLETTI, I. . A Quarta Revolução Industrial no Brasil: Arquitetura, Engenharia e a Construção Civil. *arq.urb*, n. 25, p. 1-14, 2019. DOI: 10.37916/arq.urb.vi25.11. Disponível em: <https://revistaarqurb.com.br/arqurb/article/view/11>. Acesso em: 11 jun. 2023.

NORMAN, Donald. *Design emocional*. Rio de Janeiro: Rocco, 2008. 322p.

OSOSKY, Scott. et al. Determinants of system transparency and its influence on trust in and reliance on unmanned robotic systems. *Proceedings Unmanned Systems Technology XVI*, v. 9084 n.90840, p. 1-12, 3 jun. 2014. SPIE Digital Library. <https://doi.org/10.1117/12.2050622>.

PAGE, Matthew J et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Bmj*, p. 1-7, 29 mar. 2021. *BMJ*. <http://dx.doi.org/10.1136/bmj.n71>.

PARASURAMAN, Raja; RILEY, Victor. *Humans and Automation: use, misuse, disuse, abuse*. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, v. 39, n. 2, p. 230-253, jun. 1997. SAGE Publications. <http://dx.doi.org/10.1518/00187209778543886>.

PARIKH, Viken; JOSHI, Krina. Autonomous cars using embedded systems and spherical wheels. *IJSRD - International Journal for Scientific Research & Development*. Mumbai, v. 5, n. 9, p. 146-150, Set. 2017.

PERASSO, Valeria. O que é a 4ª revolução industrial e como ela deve afetar nossas vidas. *G1/BBC Brasil, Seção Economia*. 22 out. 2016. Disponível em: <<http://g1.globo.com/economia/negocios/noti->

cia/2016/10/o-que-e-a-4a-revolucao-industrial-e-como-ela-deve-afetar-nossas-vidas.html>. Acesso em: 14 fev. 2020.

PEREIRA, S. B.; BOTELHO, R. *Design de Interação: fatores humanos e os carros autônomos*. *Design e Tecnologia*, v. 8, n. 16, p. 69-86, 31 dez. 2018.

PIAO, Jinan et al. *Public Views towards Implementation of Automated Vehicles in Urban Areas*. *Transportation Research Procedia*, v. 14, p.2168-2177, 2016. Elsevier BV. <http://dx.doi.org/10.1016/j.trpro.2016.05.232>.

PRAKKEN, Henry; SARTOR, Giovanni. *Law and logic: A review from an argumentation perspective*. *Artificial Intelligence*, v. 227, p.214-245, out. 2015. Elsevier BV. <http://dx.doi.org/10.1016/j.artint.2015.06.005>.

REDSHAW, Sarah. *In the company of cars: driving as a social and cultural practice (Human factors in road and rail transport)*. Aldershot: Ashgate Publishing Limited, 2008. 187p.

RILEY, Victor. *Operation reliance on automation: theory and data*. In: PARASURMAN, R.; MOULOUA, M., (ed.). *Automation and human performance: theory and applications*. Mahwah: Lawrence Erlbaum, 1996. p. 19-35.

SCHANK, Roger. *What is AI, anyway?*. In: PARTRIDGE, D.; WILKS, Y. *The foundations of artificial intelligence: a sourcebook*. Nova Iorque: Cambridge University Press, 1990, p. 1-13.

SCHILLER, Preston L. *Automated and Connected Vehicles: High Tech Hope or Hype?*. *World Transport Policy And Practice*, Church Stretton, v. 22, n. 3, p.28-44, out. 2016. Disponível em: <<http://worldtransportjournal.com/wp-content/uploads/2016/10/4th-Oct-opt.pdf>>. Acesso em: 21 out. 2016.

SCHWAB, Katharine. *Smart roads could protect us from self-driving cars crashes*. *Fast Company*. Fev. 2018. Disponível em: <<https://www.fastcompany.com/90160958/smart-roads-could-protect-us-from-self-driving-car-crashes>>. Acesso em: 20 out. 2019.

SCHWAB, Klaus. *A quarta revolução industrial*. São Paulo: Edipro, 2016. 159 p.

SILVA, Fernando Nunes da. *Mobilidade urbana: os desafios do futuro*. *Cadernos Metrópole*, v. 15, n. 30, p. 377-388, dez. 2013. ISSN 2236-9996. <http://dx.doi.org/10.1590/17486>.

SILVEIRA, Márcio Rogério; COCCO, Rodrigo Giraldo. *Transporte público, mobilidade e planejamento urbano: contradições essenciais*. *Estudos Avançados*. São Paulo, v. 27, n. 79, p. 41-53, 2013. <http://dx.doi.org/10.1590/S0103-40142013000300004>.

TEIXEIRA, João de Fernandes. *O que é inteligência artificial*. São Paulo: Brasiliense, 1990. 76p.

TEIXEIRA, João de Fernandes. *Mente, cérebro e cognição*. 4 ed. Petrópolis: Editora Vozes, 2011. 167p.

ZHANG, Weixiong; DECHTER, Rina; KORF, Richard E. *Heuristic search in artificial intelligence*. *Artificial Intelligence*, v. 129, n. 1-2, p.1-4, jun. 2001. Elsevier BV. [http://dx.doi.org/10.1016/S0004-3702\(01\)00111-4](http://dx.doi.org/10.1016/S0004-3702(01)00111-4).



■ Giselle Beiguelman, frames do vídeo generativo da série Flora Rebellis, versão 2024.
Imagens geradas com Inteligência Artificial (Style Gans)

■ [Acesse o vídeo clicando na imagem](#)





APRENDIZADO DE MÁQUINA E A PREVISÃO DE CURTO PRAZO PARA VELOCIDADE MÁXIMA DO VENTO

DOI: <https://doi.org/10.35699/2965-6931.2023.47653>

EDVANA DA SILVA DA SILVEIRA*; POLÍBIO ÍCARO MORO CAPO**; CARLA DE ABREU D'AQUINO***

RESUMO: Eventos de grande intensidade de ventos como passagens de frentes frias, ocorrência de ciclones, bombas, dentre outros são frequentes no sul do Brasil. A região se destaca também pela produção de energia elétrica a partir dos ventos, a qual, depende de um recurso estocástico e, portanto, sempre haverá incerteza sobre a produção final. Considerando este cenário, o presente trabalho teve como objetivo analisar a performance de modelos de machine learning para a previsão da velocidade máxima do vento com antecedência de uma hora. A regressão linear se apresentou como o método com menor capacidade de previsão. A rede neural artificial e o extreme gradient boosting (aumento de gradiente extremo, conhecido como XGBoost) apresentaram melhor desempenho em termos de métricas. Os resultados obtidos demonstraram capacidade de previsões promissoras. Esses, representam uma possibilidade de inovação na previsão de velocidades máximas de vento, principalmente, quando inserida em locais onde ocorrem diferentes fenômenos com ventos extremos.

PALAVRAS-CHAVE: previsão; velocidade máxima do vento; XGBoost; redes neurais artificiais.

Machine learning and the short-term forecast for maximum wind speed

ABSTRACT: Events of great intensity of winds, such as the passage of cold fronts, the occurrence of cyclones, bomb cyclones, among others, are frequent in southern Brazil. The region also stands out for the production of electricity from the winds, which depends on a stochastic resource and, therefore, there will always be uncertainty about the final production. Considering this scenario, the present work aimed to analyze the performance of machine learning models for predicting the maximum wind speed one hour ahead. Linear regression was the method with the lowest predictive capacity. The artificial neural network and extreme gradient boosting (XGBoost) performed better in terms of metrics. The results obtained demonstrated promising prediction capacity. These represent a possibility of innovation in the prediction of maximum wind speeds, especially when inserted in places where different phenomena with extreme winds occur.

KEYWORDS: forecast; maximum wind speed; XGBoost; artificial neural networks.

* Universidade Federal de Santa Catarina / ** Universidade Federal de Santa Catarina / *** Universidade Federal de Santa Catarina

Introdução

As mudanças climáticas têm intensificado a ocorrência de eventos com ventos extremos em todo o mundo, inclusive no Brasil. Na região Sul, verificam-se muitas frentes frias e variações bruscas de pressão, além da ocorrência de ventos intensos e eventos meteorológicos extremos, como o Furacão Catarina, ocorrido em 27 e 28 de março de 2004.

Em 2020, outro fenômeno atingiu o estado de Santa Catarina, o Ciclone Bomba, que conforme Gobato e Heidari (2020), tem como sua característica a queda repentina de pressão em seu interior, gerando ventos muito intensos. Segundo a Defesa Civil do Estado de Santa Catarina (Santa Catarina, 2021), durante o evento foram registradas velocidades acima de 130 km/h no litoral. Conforme a CELESC (Centrais Elétricas de Santa Catarina) (2020), o ciclone afetou o sistema elétrico do estado, deixando 1,5 milhão de unidades consumidoras sem energia, sendo o maior dano a rede já registrado pela companhia de energia.

Em maio de 2022, outra ocorrência de fortes rajadas de vento chamou a atenção, o Ciclone Yakecan, que conforme nota meteorológica da Defesa Civil (Santa Catarina, 2022), atingiu velocidades entre 60 a 100 km/h. Em abril de 2023, um Ciclone Bomba atinge a região do litoral sul com rajadas de vento de 79 km/h no Balneário Arroio do Silva (Santa Catarina, 2023a). No dia 13 de julho de 2023 foram emitidos diversos alertas pela Defesa Civil para vento intenso, conforme o órgão as rajadas superaram os 120 km/h em municípios do Planalto e Litoral Sul catarinense (Santa Catarina, 2023b).

As estratégias de Inteligência Artificial (IA) têm sido aplicadas em diferentes trabalhos visando a previsão da velocidade do vento, como: Peláez-Rodríguez et al. (2022); Zucatelli et al. (2019); Noorollahi, Jokar e Kalhor (2016); Samet, Reisi e Marzbani (2019); Liu et al. (2020); Navas, Prakash e Sasipraba (2020); Paula et al. (2021); Santos, Lyra e Silva Júnior (2020); Fonseca, Oliveira e Affonso (2021); Pati et al. (2023); Malakouti (2023). Entretanto, para a previsão da velocidade máxima do vento existem lacunas na literatura, principalmente, considerando a região sul do Brasil.

A previsão do recurso eólico local, segundo Camelo et al. (2018), permite o conhecimento do regime dos ventos de uma região, e dessa forma, pode auxiliar na sua exploração para geração de energia. No entanto, as previsões não são necessárias

somente para a prospecção do recurso eólico, mas também, conforme Zucatelli et al. (2019), para o planejamento, controle e monitoramento de sistemas de energia eólica. Isso permite que os operadores de parques eólicos possam tomar decisões preventivas, como desligar ou reduzir a velocidade de rotação dos aerogeradores durante condições climáticas adversas, minimizando o risco de danos e aumentando a vida útil dos equipamentos.

Com o crescimento exponencial da geração eólica e a possibilidade de geração no mar aumentam ainda mais a demanda por soluções de previsão de curto prazo das velocidades máximas de vento. Com isso, destaca-se a localização da estação meteorológica utilizada nesse estudo, que conforme Corrêa (2018) e Pires (2019), possui potencial eólico costeiro para exploração de energia eólica com futura geração *offshore*.

Nesse contexto, o presente trabalho teve por objetivo analisar o desempenho de algoritmos de *machine learning* (aprendizado de máquina) na previsão da velocidade máxima do vento com uma hora de antecedência. Assim, as investigações e os resultados contribuem para o desenvolvimento científico e tecnológico da região, bem como, podem representar uma ferramenta para auxiliar nas previsões de curto prazo, na produção de energia em um parque eólico e para a Defesa Civil na tomada de decisões sobre eventos extremos com ventos intensos.

Materiais e métodos

Os dados utilizados foram coletados em uma estação meteorológica instalada em uma zona costeira de Santa Catarina. Essa é uma região onde ocorrem diversos fenômenos com ventos intensos. A metodologia foi dividida em três etapas:

- 1) O pré-tratamento e preparação dos conjuntos de dados,
- 2) Aplicação dos modelos de *machine learning*, e
- 3) A avaliação das métricas de desempenho.

A estação meteorológica foi construída sobre uma plataforma de pesca localizada no extremo sul catarinense, no município de Balneário Arroio do Silva, conforme apresentado na Figura 1. De acordo com Pires (2019), a plataforma possui 410 m de comprimento com aproximadamente 7 m acima do nível do mar.



Figura 1 - (A) Plataforma de pesca Entremares em Balneário Arroio Silva, em vermelho a localização do laboratório, SC. (B) Laboratório de pesquisa e a torre meteorológica. Fonte: (A) Google Earth (2022); (B) MOVLIDAR-UFSC (2021), adaptado.

A Base de Observação do Oceano e Atmosfera (BOOA) é uma cooperação entre a Universidade Federal de Santa Catarina (UFSC), a plataforma de pesca Entremares e o Centro de Informações de Recursos Ambientais e de Hidrometeorologia de Santa Catarina (CIRAM/EPAGRI), a estação é identificada pelo código 1066 - Balneário Arroio do Silva – BOOA/UFSC.

Conforme Nassif (2017), a estação localiza-se na região de transição continente-oceano, com uma distância de 250 metros da linha de praia. A BOOA é um laboratório de pesquisa de 2 m², onde estão instalados diversos equipamentos, entre eles a torre meteorológica com anemômetro de copo (Movlidar-UFSC, 2021). Os dados utilizados no presente estudo foram medidos na altura de 17 m.

Pré-tratamento e preparação do conjunto de dados

Os dados utilizados foram registrados no período entre 2018 e 2021 com intervalo de 1 hora. Para o ano de 2018 foram utilizados 5419 dos 5600 observados, que compre-

endem 01/05/2018 até 31/12/2018. Em 2019 e 2020, o período de janeiro a dezembro, totalizando 8552 de 8724 e 8551 de 8768 dados, respectivamente. E para o ano de 2021, no período de janeiro a março, 2100 entre os 2155 dados observados. Em alguns meses não houve medições devido a manutenção dos equipamentos. Foram removidas as inconsistências no banco de dados em caso de falha dos sensores de medição, tais como, lacunas e variáveis com marcações '99999'.

Nos modelos de *machine learning*, normalmente, podem ser utilizadas duas formas de treinamento: não supervisionado, o qual compreende em utilizar um conjunto de variáveis e não se conhece as saídas, ou supervisionado, quando por meio de variáveis independentes, obtém-se uma variável dependente que é conhecida durante a fase de treinamento (Lee; Booth; Alam, 2005). Para o desenvolvimento do trabalho o treinamento supervisionado foi aplicado.

Os dados foram divididos em três blocos principais:

1) Treinamento: conjunto de dados utilizados para ajustar os modelos, sendo aleatoriamente selecionados 90% dos dados observados em 2018, 2019 e 2021, totalizando 14464 observações.

2) Teste: a amostra de dados usada para fornecer uma avaliação imparcial do modelo após ajustes de hiper parâmetros realizados durante o treinamento. Foram selecionados, aleatoriamente, 10% dos dados observados em 2018, 2019 e 2021, contemplando 1607 observações.

3) Validação: amostra de dados utilizada para fornecer uma avaliação imparcial do modelo final após os ajustes realizados durante o treinamento. Foram utilizados os dados do ano de 2020 com o total de 8551 dados, os quais não foram incluídos nas etapas de treinamento e teste.

Os modelos e estratégias de *machine learning* são fortemente dependentes da base de dados de treinamento. Dessa forma, a mesma contempla um período maior de observações, o que pode levar a melhores resultados na validação. Com isso, a probabilidade de ocorrerem situações já treinadas, favorece o acerto. Dado que, os modelos foram capacitados para lidar com grande variedade de situações.

O treinamento supervisionado foi aplicado considerando como variável alvo a velocidade máxima do vento 1 hora depois (objeto da previsão). Assim, verificou-se a existência das medições no instante de tempo t_0 e $t+1h$ com o objetivo de garantir que não

houvesse lacunas. Dessa forma, a base de dados foi tratada para ser um conjunto de variáveis de entrada com uma única saída, não sendo uma previsão de série temporal.

Além disso, optou-se por calcular a razão e a diferença entre as velocidades médias e máximas do vento para evitar o uso da variável alvo (velocidade máxima) durante as etapas de treinamento, teste e validação. Com isso, oito variáveis foram utilizadas como parâmetros de entrada: pressão atmosférica ao nível da estação (hPa), temperatura mínima do ar (°C), temperatura média do ar (°C), temperatura máxima do ar (°C), direção vento (°), umidade (%), razão (adimensional) e diferença entre velocidade média e máxima do vento (m/s).

Com o treinamento supervisionado, o conjunto de dados de entrada foi padronizado, conforme Equação (1).

$$z = ((x-u))/s$$

Onde u é média das variáveis de treinamento, s é o desvio-padrão, x é a variável original e z é a variável padronizada. Os dados da variável de saída (velocidade máxima do vento) estão na mesma ordem de grandeza. Portanto, não foram padronizados para utilização nos modelos de *machine learning*.

Aplicação dos modelos de machine learning

Os modelos escolhidos foram Rede Neural Artificial (RNA) e *Extreme Gradient Boosting* (XGBoost). E, para comparação de desempenho a Regressão Linear (RL) foi adotada como algoritmo base. Todos os modelos foram desenvolvidos no ambiente *Google Colaboratory* utilizando a linguagem *Python* e as bibliotecas *scikit-learn*, *keras* e *xgboost*. Cada algoritmo possui variados parâmetros de funcionamento e otimização.

Para automatizar a busca pelos melhores parâmetros foi utilizado o método *Grid Search*. O qual tem por objetivo combinar todas as possibilidades, dentro de um conjunto de parâmetros, e retornar a melhor opção (Goodfellow et al., 2016). Esse método foi realizado para a RNA e o XGBoost com validação cruzada de duas subdivisões visando poupar recursos computacionais. Já para a RL foi mantida uma configuração padrão por ser o algoritmo base para comparação de desempenho entre os modelos.

Regressão linear

A regressão linear calcula a relação entre duas ou mais variáveis, em que uma depende de outra ou outras. De maneira geral, o objetivo dessa abordagem é obter a variável dependente por meio das variáveis independentes. Nesse caso, a velocidade máxima do vento (dependente) é calculada conforme as variáveis independentes. A relação funcional de diferentes variáveis independentes com a variável dependente Y é dada pela Equação (2).

$$y = b_0 + b_1 X_1 + \dots + b_k X_k$$

Onde, y é a variável alvo, b_0 é a constante quando as variáveis independentes são 0, $b_1, b_2, \dots, b_k$ são os coeficientes e $X_1, X_2, \dots, X_k$ são os valores das variáveis independentes. A regressão linear foi considerada o modelo a ser superado, dada a sua simplicidade e ampla utilização.

Rede neural artificial

A rede neural artificial (RNA) é uma técnica computacional semelhante à característica da rede neural biológica humana (Noorollahi; Jokar; Kalhor, 2016; Ma; Feng, 2016). Uma RNA é formada por unidades de processamento, camadas (entrada, ocultas e saída), neurônios e funções de ativação (Vafaeipour et al., 2014).

Na fase de treinamento o algoritmo *backpropagation* foi utilizado. Nele, a rede é treinada com dados que consistem em um conjunto de entrada e um conjunto de saída, no qual, os pesos (conexões neurais) são ajustados para minimizar o erro entre a saída real e a prevista (Noorollahi; Jokar; Kalhor, 2016).

A configuração de uma RNA necessita de mais parâmetros além do algoritmo de treinamento. Portanto, foram utilizados: arquitetura do tipo *Multilayer Perceptron*, uma camada de entrada, uma camada oculta com função de ativação ReLu (*Rectified Linear Unit*) que consiste na mais popular para redes neurais (Oostwal; Straat; Biehl, 2021), uma camada de saída com função de ativação linear, tamanho da amostra de dados que são utilizados nos cálculos dos pesos (*batch_size* = 512) e número de épocas de 1000.

Entretanto, muitas épocas podem levar ao *overfitting* (sobre ajustado) do conjunto

de dados de treinamento, enquanto poucas podem resultar em um modelo *underfitting* (sub ajustado). A parada antecipada foi utilizada para interromper o treinamento assim que o desempenho do modelo pare de melhorar em um conjunto de dados de validação (Prechelt, 1998).

O número de neurônios na camada oculta é um fator fundamental para o desempenho da RNA. Assim, optou-se por realizar a otimização por meio do *Grid Search*, variando a quantidade de neurônios na camada oculta de 2 a 100 em intervalos de dois, totalizando 48 redes neurais avaliadas para duas validações cruzadas.

O ALGORITMO XGBOOST

O XGBoost é um algoritmo de aprendizado de máquina baseado em árvore de decisão que usa uma estrutura de aumento de gradiente e seu detalhamento matemático pode ser encontrado em Chen e Guestrin (2016).

Os hiperparâmetros determinam e afetam o processo de aprendizagem do modelo de *machine learning*, para o XGBoost estão detalhados em Chen e Guestrin (2016) e em XGBoost (2022), e estes podem ser otimizados pelas mais variadas técnicas. Os intervalos dos parâmetros tiveram como base o trabalho de Dosdo-ru e -pek (2022) e a documentação oficial disponível em XGBoost (2022). Os parâmetros dispostos na Tabela 1 correspondem a 98 modelos XGBoost treinados utilizando o *Grid Search*. O melhor modelo foi selecionado com base no menor erro calculado após a validação cruzada.

Parâmetro	Modelo proposto
<i>max_depth</i>	[3, 4, 5]
<i>learning_rate</i>	[0.01]
<i>n_estimators</i>	[1000, 2000]
<i>reg_alpha</i>	[0.067,1.1]
<i>reg_lambda</i>	[1, 1.2]
<i>Subsample</i>	[0.5]
<i>Gamma</i>	[0.8,1,5]
<i>min_child_weight</i>	[1.1,5]
<i>colsample_bytree</i>	[0.8]

Tabela 1- Lista dos parâmetros analisados utilizando o método do Grid Search.
Fonte: Organizado pelos autores, 2023.

Métricas

O desempenho e a precisão de um determinado modelo podem ser avaliados por meio de várias métricas, as quais permitem a comparação entre diferentes modelos. Conforme Sun e Wang (2018) e Kato (2016) algumas das principais métricas são: Erro Quadrático Médio (*Mean Square Error* - MSE), Equação (3), Raíz do Erro Médio Quadrático (*Root Mean Square Error* - RMSE), Equação (4), Erro Médio Absoluto (*Mean Absolute Error* - MAE), Equação (5). Chicco, Warrens e Jurman (2021) sugerem a adoção do coeficiente de determinação (R^2), Equação (6) para avaliação dos modelos de regressão.

$$MSE = \frac{1}{N} \sum_{i=1}^N e_i^2 \quad (3)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N e_i^2} \quad (4)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |e_i| \quad (5)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N e_i^2}{\sum_{i=1}^N (\bar{X} - y_{\text{observado}})^2} \quad (6)$$

Sendo $\bar{X}$ a média das observações e e_i definido pela Equação (7).

$$e_i = y_{\text{observado}} - y_{\text{previsto}} \quad (7)$$

Sendo $\bar{X}$ a média das observações e e_i definido pela Equação (7).

$$e_i = y_{(i,\text{observado})} - y_{(i,\text{previsto})} \quad (7)$$

Onde $y_{(i,\text{previsto})}$ e $y_{(i,\text{observado})}$ são os i -ésimos valores previstos e observados, respectivamente e e_i é o i -ésimo erro. Com $i = 1, \dots, N$ através de todo o conjunto de dados N . Além disso, ao utilizar o Grid Search determinou-se o menor MSE, encontrado para a base de dados de teste, como critério de seleção.

Onde $y_{(i,previsto)}$ e $y_{(i,observado)}$ são os i -ésimos valores previstos e observados, respectivamente e e_i é o i -ésimo erro. Com $i = 1,...,N$ através de todo o conjunto de dados N . Além disso, ao utilizar o Grid Search determinou-se o menor MSE, encontrado para a base de dados de teste, como critério de seleção.

As métricas expressam o desvio da previsão em relação aos dados observados. Sendo que, quando menor o valor melhor os resultados. Destaca-se que a métrica MSE, e consequentemente, o RMSE, Equações (3) e (4), atribuem maiores pesos aos maiores erros de previsão. Para o R^2 , quanto mais próximo de 1, melhor o resultado. Nesse sentido, todas as métricas apresentadas, em conjunto com a distribuição de frequência dos erros obtidos pela Equação (7), foram adotadas para um melhor entendimento dos resultados.

Resultados e discussão

Aplicando o método *Grid Search* determinou-se a melhor configuração com 56 neurônios na camada oculta da RNA. Já para o XGBoost o sumário dos parâmetros que otimizaram os resultados está apresentado na Tabela 2.

<i>Parâmetro</i>	<i>Modelo otimizado</i>
<i>max_depth</i>	4
<i>learning_rate</i>	0.01
<i>n_estimators</i>	1000
<i>reg_alpha</i>	1.1
<i>reg_lambda</i>	1
<i>Subsample</i>	0.5
<i>Gamma</i>	0.8
<i>min_child_weight</i>	1.1
<i>colsample_bytree</i>	0.8

Tabela 2 - Resultados dos parâmetros conforme o Grid Search aplicado para o modelo XGBoost. Fonte: Autores, 2023.

Com a otimização dos parâmetros foi possível realizar previsões com a base de testes e a de validação. A Tabela 3 apresenta os resultados estatísticos para ambas as bases de dados. Considerando a maior velocidade máxima medida, nenhum dos métodos reproduziu a mesma com precisão no período de testes, onde a RNA e o XGBoost subestimaram os valores e a RL superestimou (Tabela 3). Com os dados de validação, os valores da velocidade máxima previstos pela RNA e o XGBoost se aproximaram mais dos dados medidos (Linha 1), enquanto a RL não apresentou melhora.

	Teste				Validação (2020)			
		Desvio				Desvio		
Medidos	6,97	3,75	1,26	27,70	7,16	3,87	0,98	28,84
RL	7,04	3,06	-3,83	38,22	7,21	3,09	-4,49	41,87
RNA	7,08	3,32	2,18	18,58	7,23	3,41	1,69	23,12
XGBoost	7,03	3,37	1,90	18,08	7,19	3,44	1,85	20,49

Tabela 3 - Resultado da análise estatística para as bases de dados de teste/treino e validação (2020), comparando os dados reais e os três modelos de previsão escolhidos.

Fonte: Autores, 2023.

A Tabela 4 apresenta as métricas para os conjuntos de dados de teste e validação. Os valores de RMSE ficaram abaixo de 2 m/s para as RNA e XGBoost tanto no teste quanto para o ano de validação, isso demonstra que, mesmo a amostra extraída aleatoriamente da base denominada teste, os resultados foram próximos. Existem limitações ao comparar os resultados do presente trabalho com os obtidos por outros pesquisadores em diferentes regiões do mundo, uma vez que cada região tem sua particularidade geográfica e meteorológica. Contudo, os resultados obtidos por Yamaguchi e Ishihara (2021) para uma região no leste do Japão apresentaram RMSE entre 1,5 e 2 m/s para o horizonte de previsão de 1 hora.

Modelo	Teste		R ²	Validação		R ²
	MAE	RMSE		MAE	RMSE	
RL	1,66	2,17	0,652	1,71	2,29	0,649
RNA	1,13	1,60	0,802	1,17	1,71	0,805
XGB	1,11	1,60	0,812	1,15	1,69	0,810

Tabela 4 - Resultado das métricas calculadas em [m/s] para MAE e RMSE, R– adimensional. Fonte: Autores, 2023.

Ao analisar as métricas, a regressão linear se apresentou como o método com menor capacidade de previsão para as variáveis selecionadas. A RNA e o XGBoost obtiveram desempenho similar em termos de métricas, ambos melhores que a RL, o que os torna mais promissores.

A Figura 2 demonstra os histogramas com a frequência dos resíduos (Equação 7) para o ano de 2020 (base de dados de validação). Para os valores positivos os modelos realizaram previsões subestimadas em relação ao valor medido, e para os negativos a previsão foi acima do observado. Considerando ainda que, quanto maior a frequência de valores próximos de zero, melhores são os resultados das previsões.

Existem limitações nos modelos RNA e XGBoost quando se identifica que cerca de 16% da frequência dos erros é distribuída para os extremos (Figura 2b, c). Algumas hipóteses são levantadas e podem ter influenciado este comportamento: i) a região de estudo apresenta grande variação de velocidades, com médias entre 5 e 7 m/s e máximas entre 19 e 28 m/s, ii) a quantidade e qualidade dos dados, devido as falhas nas medições das séries temporais. Principalmente, pela característica da velocidade máxima de ter um comportamento de bruscas mudanças conforme Santos, Lyra e Silva Júnior (2020).

Todavia, no caso da RL (Figura 2a) a frequência de acertos representa apenas 67,60%, limitando sua aplicação. A maior frequência de erros é entre -2 a 0 e entre 0 a 2 (m/s), para os modelos de RNA (Figura 2b) e XGBoost (Figura 2c), sendo esse intervalo de 84,14% e 84,15%, respectivamente. Tal comportamento demonstra os acertos das previsões indicando, em conjunto com as métricas, que para o curto prazo esses modelos podem ser recomendados. Os resultados demonstraram que o desempenho dos algoritmos de *machine learning* analisados no presente artigo foi promissor na previsão da velocidade máxima do vento com uma hora de antecedência.

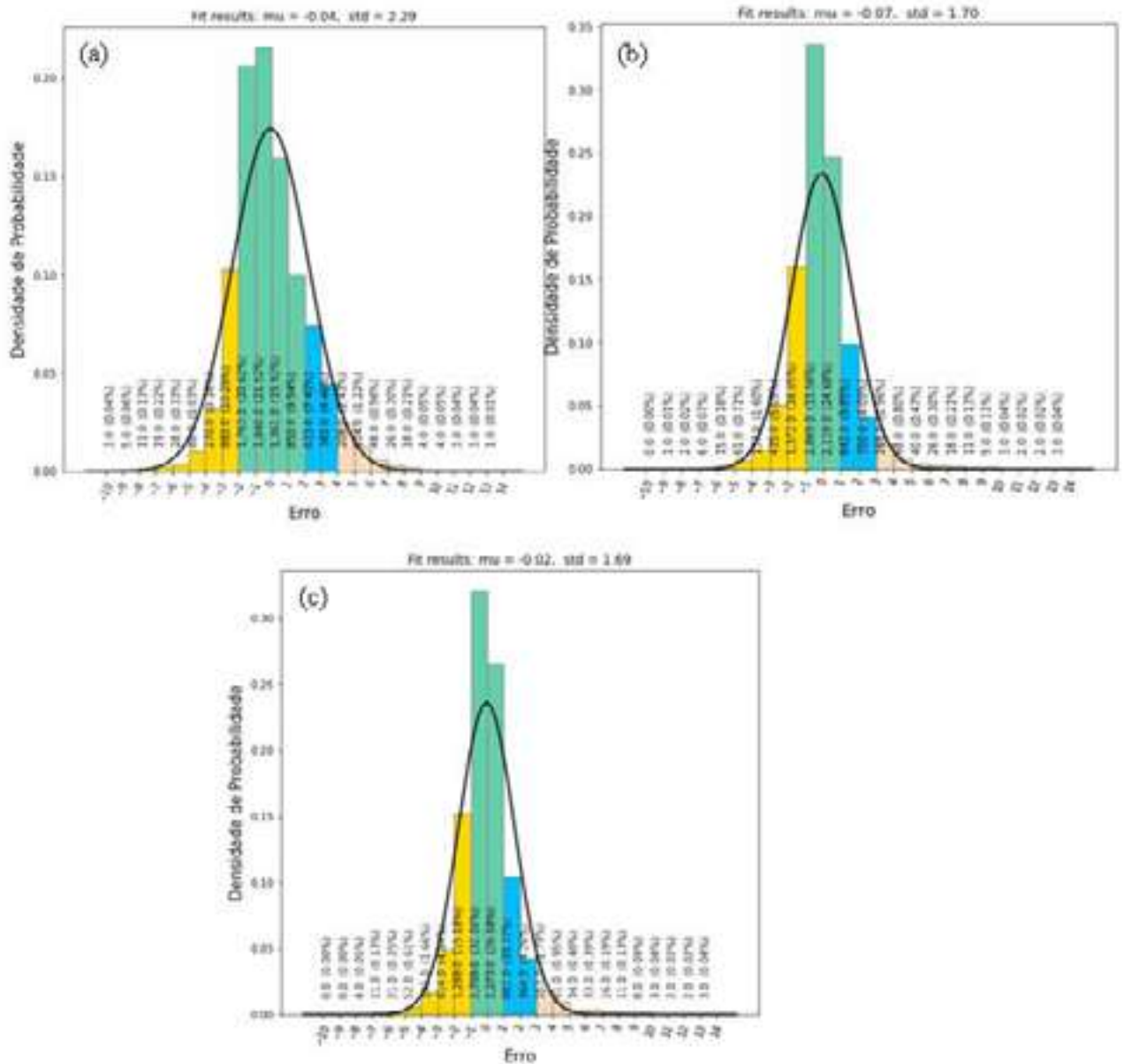


Figura 2 - Histograma com os resíduos para o ano de 2020: (a) RL, (b) RNA, (c) XGBoost. Fonte: Autores, 2023.

Considerações finais

Com a comparação dos algoritmos de *machine learning*, foi possível analisar o desempenho na previsão da velocidade máxima do vento, com uma hora de antecedência, para a região de estudo. Os resultados são relevantes, visto que eventos de vento com altas velocidades vêm se tornando cada vez mais comuns na região, causando estragos em sistemas elétricos e também podem vir a colocar em risco populações. Além disso, o crescimento exponencial da geração eólica e a possibilidade de geração no mar aumentam ainda mais a demanda por soluções de previsão de curto prazo das velocidades máximas de vento.

Dessa forma, destaca-se a relevância do presente trabalho o qual utilizou dados inéditos de uma estação meteorológica em zona costeira que sofre com a influência de fenômenos meteorológicos de grande intensidade de ventos. E, também, da possibilidade de futura geração *offshore* no local. Espera-se que as contribuições e o pioneirismo do estudo norteiem para a continuação e aperfeiçoamento dos algoritmos e métodos de *machine learning* focando na previsão da velocidade máxima dos ventos. Bem como, a reprodução da metodologia em outras localidades do Brasil para comparação de desempenho.

Referências

- CAMELO, H. do N.; LUCIO, P. S.; LEAL JUNIOR, J. B. V.; CARVALHO, P. C. M. de. Proposta para Previsão de Velocidade do Vento Através de Modelagem Híbrida Elaborada a Partir dos Modelos ARIMAX e RNA. *Revista Brasileira de Meteorologia*, v. 33, p. 115–129, mar. 2018. <http://dx.doi.org/10.1590/0102-7786331005>.
- CELESC. Centrais Elétricas de Santa Catarina. Ciclone: Celesc restabelece sistema para cerca de 1 milhão de unidades. 2020. Disponível em: <https://www.celesc.com.br/listagem-noticias/ciclone-celesc-restabelece-sistema-para-cerca-de-1-milhao-de-unidades>. Acesso em: 23 abr. 2023.
- CHEN, T.; GUESTRIN, C. XGBoost: A Scalable Tree Boosting System. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco California USA: ACM, 2016, p. 785–794. <http://dx.doi.org/10.1145/2939672.2939785>.
- CHICCO, D.; WARRENS, M. J.; JURMAN, G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, v. 7, p. e623, 5 jul. 2021. <http://dx.doi.org/10.7717/peerj-cs.623>.

CORRÊA, A. G. *Climatologia dos Ventos e Potencial Eólico Offshore de Santa Catarina*. 2018. 85 f. Dissertação (Mestrado), Curso de Oceanografia, Universidade Federal de Santa Catarina, Florianópolis, 2018.

DOSDOBRU, A. T.; BPEK, A. B. Hybrid boosting algorithms and artificial neural network for wind speed prediction. *International Journal of Hydrogen Energy*, v. 47, n. 3, p. 1449–1460, jan. 2022. <http://dx.doi.org/10.1016/j.ijhydene.2021.10.154>.

FONSECA, S. B.; OLIVEIRA, R. C. L.; AFFONSO, C. M. Short-term Wind Speed Forecasting using Machine Learning Algorithms. 28 jun. 2021. 2021 IEEE Madrid PowerTech, Madrid, Spain: IEEE, 28 jun. 2021. p. 1–6. <http://dx.doi.org/10.1109/powertech46648.2021.9494848>.

GOBATO, R.; HEIDARI, A. Cyclone Bomb Hits Southern Brazil in 2020. *Journal Of Atmospheric Science Research*, v. 3, n. 3, 4 ago. 2020. <http://dx.doi.org/10.30564/jasr.v3i3.2163>.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep learning*. Cambridge, Massachusetts: The MIT Press, 2016.

Google Earth website. Plataforma de pesca Entremares em Balneário Arroio do Silva. 2022. Disponível em: <https://earth.google.com/web/@-28.96237375,-49.37920088,-0.18856225a,806.36277371d,35y,oh,ot,or>. Acesso em: 03 abr. 2023

KATO, T. Prediction of photovoltaic power generation output and network operation. *Integration Of Distributed Energy Resources In Power Systems*, [S.L.], p. 77-108, 2016. Elsevier. <http://dx.doi.org/10.1016/b978-0-12-803212-1.00004-0>.

LEE, K; BOOTH, D; ALAM, P. A comparison of supervised and unsupervised neural networks in predicting bankruptcy of Korean firms. *Expert Systems With Applications*, v. 29, n. 1, p. 1-16, jul. 2005. Elsevier BV. <http://dx.doi.org/10.1016/j.eswa.2005.01.004>.

LIU, X.; ZHANG, H.; KONG, X.; LEE, K. Y. Wind speed forecasting using deep neural network with feature selection. *Neurocomputing*, v. 397, p. 393–403, jul. 2020. <http://dx.doi.org/10.1016/j.neucom.2019.08.108>.

MA, Y.; FENG, S. Prediction and Forecast of Wind Power. *Large-Scale Wind Power Grid Integration*, [S.L.], p. 185-209, 2016. Elsevier. <http://dx.doi.org/10.1016/b978-0-12-849895-8.00005-1>.

MALAKOUTI, S. M. Estimating the output power and wind speed with ML methods: a case study in Texas. *Case Studies in Chemical and Environmental Engineering*, v. 7, p. 100324, jun. 2023. Elsevier BV. <http://dx.doi.org/10.1016/j.cscee.2023.100324>.

MOVLIDAR-UFSC. Base de Observação do Oceano e Atmosfera. 2021. Disponível em: <https://movlidar.paginas.ufsc.br/estruturas/booa>. Acesso em: 13 abr. 2023

NASSIF, F. B. *A tecnologia LIDAR aplicada a medições eólicas sobre corpos hídricos e oceano*. 2017. 111 f. Dissertação (Mestrado) - Curso de Programa de Pós-Graduação em Engenharia Mecânica, Centro Tecnológico, Universidade Federal de Santa Catarina, Florianópolis, 2017.

NAVAS, R. K. B.; PRAKASH, S.; SASIPRABA, T. Artificial Neural Network based computing model for wind speed prediction: A case study of Coimbatore, Tamil Nadu, India. *Physica A: Statistical Mechanics and its Applications*, v. 542, p. 123383, mar. 2020. <http://dx.doi.org/10.1016/j.physa.2019.123383>.

NOOROLLAHI, Y.; JOKAR, M. A.; KALHOR, A. Using artificial neural networks for temporal and

spatial wind speed forecasting in Iran. *Energy Conversion and Management*, v. 115, p. 17–25, mai. 2016. <http://dx.doi.org/10.1016/j.enconman.2016.02.041>.

OOSTWAL, E.; STRAAT, M.; BIEHL, M. Hidden unit specialization in layered neural networks: ReLu vs. sigmoidal activation. *Physica A: Statistical Mechanics and its Applications*, v. 564, p. 125517, fev. 2021. <http://dx.doi.org/10.1016/j.physa.2020.125517>.

PATI, N.; GOURISARIA, M. K.; DAS, H.; BANIK, D. Wind Speed Prediction using Machine Learning Techniques. 28 abr. 2023. 2023 11th International Conference on Emerging Trends in Engineering & Technology - Signal and Information Processing (ICETET - SIP), Nagpur, India: IEEE, 28 abr. 2023. p. 1–6. <http://dx.doi.org/10.1109/icetet-sip58143.2023.10151597>.

PAULA, M.; COLNAGO, M.; FIDALGO, J. N.; CASACA, W. Predicting Long-Term Wind Speed in Wind Farms of Northeast Brazil: A Comparative Analysis Through Machine Learning Models. *IEEE Latin America Transactions*, [S. l.], v. 18, n. 11, p. 2011–2018, 2021.

PELÁEZ-RODRÍGUEZ, C.; PÉREZ-ARACIL, J.; FISTER, D.; PRIETO-GODINO, L.; DEO, R. C.; SALCEDO-SANZ, S. A hierarchical classification/regression algorithm for improving extreme wind speed events prediction. *Renewable Energy*, v. 201, p. 157–178, dez. 2022. <http://dx.doi.org/10.1016/j.renene.2022.11.042>.

PIRES, C. H. M. Avaliação dos recursos eólicos com um LIDAR instalado em uma plataforma costeira do sul do Brasil. 2019. 98 f. Dissertação (Mestrado) - Curso de Oceanografia, Universidade Federal de Santa Catarina, Florianópolis, 2019.

PRECHELT, L. Automatic early stopping using cross validation: quantifying the criteria. *Neural Networks*, v. 11, n. 4, p. 761–767, jun. 1998. [http://dx.doi.org/10.1016/s0893-6080\(98\)00010-0](http://dx.doi.org/10.1016/s0893-6080(98)00010-0).

SAMET, H.; REISI, M.; MARZBANI, F. Evaluation of neural network-based methodologies for wind speed forecasting. *Computers & Electrical Engineering*, v. 78, p. 356–372, set. 2019. <http://dx.doi.org/10.1016/j.compeleceng.2019.07.024>.

SANTA CATARINA. Defesa Civil realiza “Live” sobre o “Ciclone Bomba” registrado em 2020. Defesa Civil de Santa Catarina, 2021. Disponível em: <https://www.defesacivil.sc.gov.br/noticias/defesa-civil-realiza-live-sobre-o-ciclone-bomba-registrado-em-2020/>. Acesso em: 06 abr. 2023.

SANTA CATARINA. Nota Meteorológica da Defesa Civil /SC - 16/05 – possível Formação da Tempestade Subtropical Yakecan nesta terça-feira (17). Defesa Civil de Santa Catarina, 2022. Disponível em: <http://www.defesacivil.sc.gov.br/noticias/notas-oficiais/nota-meteorologica-dc-sc-16-05-possivel-formacao-da-tempestade-subtropical-yakecan-nesta-terca-feira-17/> Acesso em: 10 abr. 2023.

SANTA CATARINA. Nota meteorológica da Defesa Civil /SC - 03/04 – Ciclone bomba provoca ressaca e alagamentos costeiros em Santa Catarina. Defesa Civil de Santa Catarina, 2023a. Disponível em: <https://www.defesacivil.sc.gov.br/noticias/nota-meteorologica-dc-sc-03-04-ciclone-bomba-provoca-ressaca-e-alagamentos-costeiros-em-santa-catarina/>. Acesso em: 04 out. 2023.

SANTA CATARINA. Monitoramento Meteorológico Defesa Civil /SC 13/07 10:40. Defesa Civil de Santa Catarina, 2023b. Disponível em: <https://www.defesacivil.sc.gov.br/previsao-do-tempo/monitoramento-meteorologico-dc-sc-13-07-1020/>. Acesso em: 04 out. 2023.

SANTOS, I. G. DA S.; LYRA, R. F. DA F.; SILVA JÚNIOR, R. S. DA. Comparativo de Prognósticos da

Velocidade do Vento Utilizando Modelo WRF e Rede Neural Artificial. Revista Brasileira de Meteorologia, v. 35, n. spe, p. 1017–1027, dez. 2020. <http://dx.doi.org/10.1590/0102-77863550103>.

SUN, W.; WANG, Y. *Short-term wind speed forecasting based on fast ensemble empirical mode decomposition, phase space reconstruction, sample entropy and improved back-propagation neural network. Energy Conversion and Management*, v. 157, p. 1–12, fev. 2018. <http://dx.doi.org/10.1016/j.enconman.2017.11.067>.

VAFAEIPOUR, M.; RAHBARI, O.; ROSEN, M. A.; FAZELPOUR, F.; ANSARIRAD, P. *Application of sliding window technique for prediction of wind velocity time series. International Journal of Energy and Environmental Engineering*, v. 5, n. 2–3, p. 105, 2014. <http://dx.doi.org/10.1007/s40095-014-0105-5>.

XGBOOST. *XGBoost Parameters*. 2022. Disponível em: <https://xgboost.readthedocs.io/en/latest/parameter.html#general-parameters>. Acesso em: 20 abr. 2023.

YAMAGUCHI, A.; ISHIHARA, T. *Maximum Instantaneous Wind Speed Forecasting and Performance Evaluation by Using Numerical Weather Prediction and On-Site Measurement. Atmosphere*, v. 12, n. 3, p. 316, 28 fev. 2021. <http://dx.doi.org/10.3390/atmos12030316>.

ZUCATELLI, P.J.; NASCIMENTO, E.G.s.; AYLAS, G.y.R.; SOUZA, N.B.P.; KITAGAWA, Y.K.L.; SANTOS, A.A.B.; ARCE, A.M.G.; MOREIRA, D.M. *Short-term wind speed forecasting in Uruguay using computational intelligence. Heliyon*, [S.L.], v. 5, n. 5, e01664, maio 2019. <http://dx.doi.org/10.1016/j.heliyon.2019.e01664>.



■ Giselle Beiguelman, frames do vídeo generativo da série Flora Rebellis, versão 2024.
Imagens geradas com Inteligência Artificial (Style Gans)

■ [Acesse o vídeo clicando na imagem](#)





ESTUDO COMPARATIVO DE MODELOS PARA PREVISÃO DA MORTALIDADE DA SRAG POR COVID-19 NO BRASIL

DOI: <https://doi.org/10.35699/2965-6931.2023.47653>

ANDRÉ VINÍCIUS MENDES BARROS*; VICTOR HUGO DANTAS GUIMARÃES** LUCIANA BALIEIRO COSME***

RESUMO: A síndrome respiratória aguda grave 2 (SARS-CoV-2) compreende uma das complicações desencadeadas pelo novo coronavírus. O presente trabalho tem como objetivo propor uma comparação entre dois modelos baseados em aprendizagem de máquina em diferentes contextos para prever a mortalidade nos casos da Síndrome Respiratória Aguda Grave (SRAG) pelo coronavírus 2019, COVID-19. Os dados utilizados estão disponíveis na plataforma DataSUS, e compreendem o período de janeiro de 2021 a dezembro de 2022. Consequentemente, realizou-se uma análise estatística descritiva, seleção de variáveis e, por fim, a elaboração de dois modelos, um antes do marco da segunda dose de vacinação para COVID-19 e outro depois. Em relação às métricas, o modelo 1 apresentou uma acurácia de 71,8%, enquanto o modelo 2 obteve uma acurácia de 80%, contribuindo portanto no processo de tomada de decisão para o enfrentamento da doença.

PALAVRAS-CHAVE: modelos preditivos; regressão logística; COVID-19

Comparative study of models for predicting COVID-19-associated Severe Acute Respiratory Syndrome (SARS) mortality in Brazil

ABSTRACT: Severe Acute Respiratory Syndrome 2 (SARS-CoV-2) comprises one of the complications triggered by the new coronavirus. This study aims to propose a comparison between two machine learning-based models in different contexts to predict mortality in cases of Severe Acute Respiratory Syndrome (SARS) caused by the 2019 coronavirus, COVID-19. The data used are available on the DataSUS platform and cover the period from January 2021 to December 2022. Consequently, a descriptive statistical analysis, variable selection, and the development of two models were carried out, one before the milestone of the second dose of COVID-19 vaccination and another after. Regarding the metrics, model 1 presented an accuracy of 71.8%, while model 2 achieved an accuracy of 80%, thus contributing to the decision-making process for tackling the disease.

KEYWORDS: predictive models; logistic regression; COVID-19.

* Instituto Federal do Norte de Minas Gerais/ ** Universidade Estadual de Montes Claros /

*** Instituto Federal do Norte de Minas Gerais

Introdução

Descoberto em dezembro de 2019 na cidade de Wuhan (China), o vírus de RNA envelopado de fita simples, pertencente ao gênero *Betacoronavirus* e família *Coronaviridae*, e nomeado como SARS-CoV-2 (ALEXANDER E. GORBALENYA, 2020; de GROOT R.J., 2020; LU *et al.*, 2020), se espalhou rapidamente pelo mundo e acabou se transformando em um grave problema de saúde pública global, segundo HUANG *et al.* (2020). Por exemplo, comprovando as proporções do problema, dados mostraram que, em maio de 2022, mais de 528 milhões de pessoas em todo o mundo já haviam sido afetadas pela doença e mais de 6 milhões de óbitos haviam sido registrados. Em maio de 2023, porém, a Organização Mundial da Saúde (OMS) decretou o fim da Emergência de Saúde Pública de Importância Internacional referente ao SARS-CoV-2, mas, ainda assim, estudos sobre o tema são cada vez mais relevantes para que a doença possa ser melhor compreendida e seus efeitos negativos cada vez mais mitigados. Sob essa ótica, dados do DataSUS mostram que, nos casos de SARS-CoV-2 (COVID-19), a síndrome respiratória aguda é ainda imperativa para a mortalidade nos casos agravados (LI *et al.*, 2020). Além disso, estudos indicam uma taxa de mortalidade maior quando têm-se complicações como diabetes, doenças cardiovasculares, obesidade e doenças respiratórias prévias associadas (CUCINOTTA & VANELLI, 2020; FARIA, 2021; ZHOU *et al.*, 2021). Dentro dessa população, os idosos se destacam como um grupo de alto risco devido à subnutrição, baixa imunológica e diversas doenças crônicas associadas, que eles comumente podem apresentar. Desta forma, os dados disponíveis também apontam para uma importante questão de saúde pública no Brasil, tendo em vista a transição demográfica marcada pelo aumento de pessoas idosas, que corresponde a cerca de 14% (30,9 milhões) da população no país (KÜCHEMANN, 2012; CARVALHO & RODRÍGUEZ-WONG, 2008). Além disso, durante 2022, tornou-se evidente uma significativa redução no número de casos, especialmente após o alcance do marco de 70% da população brasileira vacinada com a segunda dose da vacina contra a COVID-19, registrado em 01/02/2022. Com o objetivo de analisar o

impacto desse cenário pré e pós-vacinação, foram desenvolvidos dois modelos distintos. O primeiro modelo abrangeu o período de 01/01/2021 a 01/02/2022, enquanto o segundo modelo englobou o período de 01/02/2022 a 31/12/2022. Esses modelos foram criados para investigar a influência da vacinação COVID-19 nos resultados e fornecer percepções sobre a eficácia das medidas de imunização em relação à SARS-CoV-2 e seus desdobramentos na saúde pública.

Neste sentido, o presente estudo tem como objetivo realizar uma comparação entre dois modelos baseados em aprendizado de máquina que utilizam dados socio-demográficos e fisiológicos para prever a mortalidade de pacientes com SRAG por COVID-19. O primeiro modelo é construído com dados antes do marco de 70% de vacinação da COVID-19, estimado em 01/02/2022, enquanto o segundo modelo é construído com dados após essa data.

Referencial teórico

O termo aprendizado de máquina é caracterizado como a área que estuda como usar computadores para simular atividades de aprendizado humano e estudar métodos de auto-aperfeiçoamento de computadores para obter novos conhecimentos e novas habilidades, identificar o conhecimento existente e melhorar continuamente o desempenho e sua realização, segundo WANG *et al.* (2009). Para esse trabalho, o aprendizado de máquina pode ser entendido como a área que reúne algoritmos que são capazes de simular atividades do aprendizado humano. Essa área possui três diferentes tipos de aprendizado: o aprendizado supervisionado, não supervisionado e por reforço.

O aprendizado supervisionado é um ramo do aprendizado de máquina, em que se tem como características o conhecimento dos vetores de entrada, valores que serão utilizados como entrada do modelo; e dos vetores de saída (BISHOP, 2006), valores que serão utilizados como saída do modelo, o que se deseja prever. Esse tipo de aprendizado tem por intuito ajustar um modelo que relacione a resposta (vetor de saída) aos preditores (vetor de entrada). O objetivo é prever com precisão a resposta para observações futuras ou entender melhor a relação entre a resposta e os preditores (JAMES *et al.*, 2013). Essa previsão ainda pode ser para: reconhecer uma determinada categoria

de elementos em um conjunto de dados (algoritmos de classificação) ou prever um conjunto de valores nominais (algoritmos de regressão).

Formalmente, um problema de classificação possui: um conjunto de i entradas $X = \{X_1, X_2, \dots, X_i\}$, um conjunto de j classes $C = \{c_1, c_2, \dots, c_j\}$, sendo que, para cada conjunto X_i , uma ou mais classes c_i serão seus rótulos; e por último, um conjunto de n vetores de treinamento $T = \{(x_1, \{c_1, \dots, c_j\}), \dots, (x_n, \{c_1, \dots, c_n\})\}$. Com esses elementos, já é possível realizar o treinamento do modelo de classificação, por meio do uso do conjunto de vetores de treinamento que são inseridos para aprender ou estimar os parâmetros desconhecidos do modelo (DUDA *et al.*, 2000). Ao identificar um conjunto de novas entradas, o modelo pode ser capaz de classificar a qual classe cada elemento do conjunto de entradas pertence. Esse estudo utiliza dois modelos de aprendizado de máquina, a regressão logística e a floresta aleatória.

Regressão logística

Método paramétrico originado a partir da estatística e popular na área da saúde (DREISEITL, 2002). Esse método é considerado paramétrico pois resume os dados com um conjunto de parâmetros de tamanho fixo, ou seja, caso tenha N variáveis $X_1, X_2, X_3, \dots, X_n$, o modelo terá obrigatoriamente $N + 1$ parâmetros, conforme explicado por NORVIG (2010).

A regressão logística tem como premissa ajustar os parâmetros do modelo de forma a minimizar a perda da soma dos resíduos dos quadrados no conjunto de dados (NORVIG, 2010; DAVID DIE, 2022). A soma dos resíduos dos quadrados é dada pela Equação 1 e se refere à soma das diferenças entre o valor real (y_i) e o previsto ($\hat{y}_i$) ao quadrado. A vantagem de avaliar a regressão utilizando esse método é poder mensurar quão longe a resposta está do valor esperado.

$$SSE = \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (1)$$

Para a previsão, a regressão logística faz o uso de um valor limiar entre 0 e 1 que,

para valores que sejam maiores ou iguais a esse limiar, o modelo caracteriza como uma dada classe ou para valores abaixo desse limiar, como outra classe. Cria-se dessa forma, um classificador linear (NORVIG, 2010). Ou seja, o que foi gerado pela função linear dada na Equação 2 é passado na função de limiar para verificar a qual classe aquele dado pertence. Na mesma equação o termo z é a própria função linear originada a partir do conjunto de dados. Assim, ao aplicá-la na função de limiar descrita, cria-se um classificador linear, sendo possível, portanto, realizar a classificação dos dados.

$$\text{Logística}(z) = 1 / (1 + e^{(-z)}) \quad (2)$$

Dentre os atributos que a função de regressão logística traz consigo, pode-se citar os valores dos coeficientes da regressão e a razão de chances (em inglês, *odds ratio*, OR). Os coeficientes da regressão são dados por números que representam o quanto uma dada variável será multiplicada pelo valor presente naquela variável para obtenção da previsão, possibilitando que um evento tenha probabilidade maior ou menor de ocorrer. Na Equação 3, β_0 e β_1 são os coeficientes das variáveis x_1 e x_2 , respectivamente. Esses coeficientes são estimados, na maioria das vezes, pelo método de máxima verossimilhança, que possui a fórmula descrita na Equação 4. Segundo SUBHASH (1995), esse método busca estimar os parâmetros β_0 e β_1 , de modo que, ao inserir essas estimativas no modelo para y , dado na Equação 3, produza um número próximo do objetivo real.

$$y = 1 / (1 + e^{(-\beta_0 - \beta_1 x_1 - \beta_2 x_2)}) \quad (3)$$

$$L(\beta_0, \beta_1) = \sum_{i=1}^n (y_i \ln(\beta_0 + \beta_1 x_i) + (1 - y_i) \ln(1 - \beta_0 - \beta_1 x_i)) \quad (4)$$

Conforme representado na Equação 5, a OR é uma medida de associação que traz consigo a razão entre duas probabilidades (AGUIAR & NUNES, 2013), a probabilidade de ocorrência de um evento (óbito) pela probabilidade de não ocorrência (cura). Por exemplo, suponha-se que a variável diabetes tenha um OR = 2, ou seja, ao dividir a probabilidade de uma pessoa com diabetes morrer e a probabilidade de uma pessoa sem diabetes morrer, tem como resultado 2. Considera-se que, esse paciente que tem diabetes, possui 2 vezes mais chances de ir a óbito.

$$OR = (-\log \log p_{-1} / (1 - p_{-1})) / (-\log \log p_{-2} / (1 - p_{-2})) \quad (5)$$

Diferente da regressão logística que se baseia na estatística, a floresta aleatória (do inglês, random forest) é um método não-paramétrico que se baseia em múltiplas árvores de decisão. Ser considerado não-paramétrico se deve ao fato do modelo não ter suposições sobre a distribuição do espaço e a estrutura do classificador, ou seja, não se tem uma quantidade de parâmetros fixos como a regressão logística (MAIMON & ROKACH, 2010). NORVIG (2010) descreve uma árvore de decisão como sendo uma representação para uma função que toma como entrada um vetor de valores de atributos e retorna uma decisão - de acordo com ele o retorno é de um único valor de saída. Em sua estrutura, a árvore de decisão é constituída por nós e ramificações, o primeiro nó recebe o nome de raiz e os últimos nós da parte baixa da árvore são chamados de nós folhas. Esse método funciona a partir das decisões tomadas em cada nó, particionando a entrada de modo que, ao final do particionamento, os nós folhas contenham elementos de uma única classe (RATSCH, 2004). Essa árvore tem como estrutura vários nós ligados de forma hierárquica: o nó raiz pode possuir filhos e cada um desses filhos outros filhos, assim sucessivamente. Para realizar a transição de um nó para o outro, deve ser verificada qual a decisão tomada naquele nó.

Como vantagens, podem ser citadas a sua fácil compreensão, possibilidade de ter dados de entrada nominais, numéricos e faltantes (MAIMON & ROKACH, 2010). E como desvantagens, a super adaptação (do inglês, overfitting), que é quando uma árvore se adapta muito a um determinado padrão durante o treinamento (NORVIG, 2010).

O algoritmo da floresta aleatória segue o mesmo princípio de um algoritmo baseado em árvore de decisão, mas o modelo conta com diversas árvores construídas aleatoriamente de maneira a prover diversidade ao modelo (LIMA et al., 2021). Cada árvore, ao receber um subconjunto aleatório do conjunto de dados, construirá a árvore de decisão e será capaz de classificar um dado conjunto de entrada. Como são múltiplas árvores de classificação, ao final da criação das árvores, a floresta adotará alguma estratégia para um problema em dados discretos, como por exemplo, escolher a árvore com mais votos e utilizá-la para classificar o conjunto de dados inserido (SENA, 2021; BREIMAN, 2001). Além do fato de ser um algoritmo de aprendizado supervisionado,

a floresta aleatória pode ser utilizada como técnica para seleção de variáveis e, consequentemente, para redução da dimensionalidade do problema de modo a identificar quais variáveis foram mais importantes para a classificação (SENA, 2021). Na subseção seguinte serão apresentadas duas técnicas para seleção de variáveis, a seleção passo a passo para frente e o método da importância das variáveis (em inglês, *feature importances*) pela floresta aleatória, que busca classificar as variáveis por sua importância.

Métricas de avaliação

A primeira técnica utilizada para auxiliar no processo de avaliação de um modelo é a validação cruzada. FARIA (2021) define a validação cruzada como sendo uma ferramenta estatística utilizada para avaliar a capacidade de generalização de um modelo a partir de um determinado conjunto de dados não utilizado para obter as estimativas do desempenho do referido modelo. Sendo assim, com essa técnica é possível obter uma estimativa mais precisa do modelo e, caso se tenha mais de um, verificar aquele que se generalizou melhor aos dados (NORVIG, 2010), apresentando-se como sugestão para a avaliação do desempenho de cada modelo. No método de validação cruzada com *k* repetições, a base de dados é dividida em *k* blocos iguais e, a cada rodada de treinamento, as informações referentes ao treino e teste mudam. Ao final dessas *k* repetições é realizada uma média do resultado de cada repetição do modelo. Um exemplo de resultado seria a quantidade de erros presentes nos dados de teste sendo possível reduzir a probabilidade de se ter pontuações muito boas ou muito ruins que não condizem com o modelo.

Para comparar os resultados de cada classe, têm-se as métricas de precisão, acurácia, sensibilidade e especificidade, que são bastante utilizadas em trabalhos relacionados com o tema. Já as métricas de desempenho se referem a quantidade de acertos a partir de um determinado ponto de vista. Para entender melhor sobre as métricas de desempenho de um modelo de aprendizado de máquina é preciso explicar os conceitos de verdadeiro positivo, falso positivo, verdadeiro negativo e falso negativo, a saber, segundo PROVOST (1998) e FRANCESCHI (2019):

Verdadeiro positivo (VP): número de óbitos que foram preditos como óbito;
Falso positivo (FP): número de curados que foram preditos como óbito;
Verdadeiros negativos (VN): número de curados que foram preditos como curados;
Falso negativo (FN): número de óbitos que foram preditos como curado.

Óbitos		Reais	
		Curados	
Preditos	Óbitos	VP	FP
	Curados	FN	VN

Tabela 1 - Matriz de confusão para o modelo de classificação. Fonte: Elaborado pelo autor (2023)

A matriz de confusão de um modelo de classificação, representada na Tabela 1, é um instrumento importante pois apresenta diversas estatísticas originadas a partir de si, como por exemplo a acurácia, precisão e especificidade, possuindo portanto mais informações sobre o desempenho do modelo. Diante do contexto abordado, duas métricas podem ser priorizadas: a acurácia e a precisão, ambas utilizadas no aprendizado de máquina e descritas por PROVOST (1998).

A acurácia, Equação 6, busca um equilíbrio entre os acertos dentro de duas classes (curados ou óbitos) apresentando o percentual de acertos sobre a variabilidade das predições realizadas. A precisão ou especificidade (em inglês, precision), Equação 7, informa quanto o modelo acertou da classe positiva ao analisar os valores preditos e a sensibilidade, dada na Equação 8, informa o quanto o modelo acertou os valores reais da classe positiva. Nesse último caso, busca-se analisar mais detalhadamente os valores reais da classe positiva, classe que analisa os óbitos, pois é mais aceitável errar o desfecho de um paciente curado, que foi predito como óbito, do que o contrário.

acurácia= (VP+VN)/(VP+FP+FN+VN) (6)

especificidade= VP/(VP+FP) (7)

sensibilidade= VP/(VP+FN) (8)

E a última métrica será a curva de característica de operação do receptor (AUCROC). Inicialmente, a curva ROC se baseia no cálculo da taxa de verdadeiros positivos (sensibilidade) e taxa de falsos positivos, conforme a Equação 8 e Equação 9 respectivamente.

te. Essa curva se apresenta no gráfico em que o eixo X corresponde a taxa de falsos positivos e, o eixo Y corresponde a taxa de verdadeiros positivos, uma observação é que ambos os eixos possuem valor mínimo 0 e máximo 1. Com essa curva é possível analisar o desempenho de vários modelos com diferentes performances nas métricas apresentadas nessa seção e, assim, tomar a decisão de qual modelo escolher.

$$\text{taxa de falsos positivos} = 1 - \text{especificidade} \quad (9)$$

Balanceamento de dados

Ao analisar o número de elementos em cada classe do trabalho, curados e óbitos, depara-se com um desbalanceamento das classes de maneira que o número de curados é muito maior que o número de óbitos, e consequentemente, uma classe (óbito) pode estar sub-representada no conjunto de dados (HERRERA, 2018). No intuito de equilibrar as classes para que os modelos a reconheçam melhor, pode-se fazer o uso por exemplo de técnicas de balanceamento de classes baseadas em aprendizado sensível ao custo (LUO & PAN, 2019).

Segundo LUO & PAN (2019), essa estratégia aborda o problema de desequilíbrio no nível de dados e no nível algorítmico, no qual um alto custo de classificação incorreta da classe minoritária é atribuído e o custo geral é minimizado. Sendo assim, o objetivo é ajustar os pesos de cada classe de modo que o modelo dê uma importância maior, as instâncias das classes minoritárias e o contrário para as classes majoritárias. Neste trabalho, utiliza-se essa abordagem a nível de dados devido à simplicidade e já estar incorporado à biblioteca utilizada. No algoritmo de regressão logística disponível na biblioteca scikit-learn por exemplo, o balanceamento das classes pode ser feito automaticamente utilizando o parâmetro *class_weight* com o valor 'balanced'. Por meio dessa configuração, o algoritmo ajusta os pesos automaticamente invertendo proporcionalmente a frequência de cada classe, conforme pode ser visto na Equação 10.

$$\text{peso}_{\text{classe } i} = (\text{Tamanho do conjunto de dados}) / (\text{Quantidade de classes} * \text{Quantidade de amostras da classe } i) \quad (10)$$

Metodologia

Como tecnologias para o desenvolvimento, fez-se o uso da linguagem Python (<https://www.python.org/>) na versão 3.10.12, e das bibliotecas da própria linguagem como: Pandas (<https://pandas.pydata.org/>), para manipulação e análise dos dados; Numpy (<https://numpy.org/pt/>), para manipulação de operações aritméticas; scikit-learn (<https://scikit-learn.org/stable/>), para criação de modelos matemáticos e uso de métricas de avaliação dos modelos, e matplotlib (<https://matplotlib.org/>) e seaborn (<https://seaborn.pydata.org/>), para criação de gráficos. A base de dados para esse trabalho é de domínio público, sendo assim possuem a licença Creative Commons Atribuição 4.0 Internacional (CC BY 4.0) e são disponibilizados pelo openDataSUS (<https://opendatasus.saude.gov.br/dataset/srag-2020>), site que disponibiliza dados acerca da saúde no Brasil. Os dados compreendem o período de 01/01/2021 à 31/12/2022, totalizando 3.476.568 registros e uma prévia pode ser visualizada na Figura 1. Devido ao tema do trabalho, foram filtrados apenas os casos que tiveram como diagnóstico a SRAG por COVID-19 e selecionados os pacientes que tiveram desfecho, ou seja, foram curados ou morreram. Nesse trabalho, optou-se por não imputar valores, ou seja, caso a variável for preenchida pelo valor “Ignorado” ou estiver nula, a linha toda daquela informação foi retirada. Como consequência, obteve-se um menor número de informações na base de dados.

CS_SEXO	NÚ_IDADE_H	CS_RACA	CS_ESCOL_N	FEBRE	TOSSE	GARGANTA	DISPNEIA	DESC_RESP	SATURACAO
F	56	1.0	9.0	1.0	1.0	1.0	2.0	1.0	2.0
F	67	1.0	9.0	1.0	1.0	2.0	1.0	1.0	1.0
M	68	1.0	9.0	1.0	1.0	2.0	2.0	2.0	2.0
F	42	2.0	9.0	1.0	1.0	2.0	1.0	1.0	1.0
M	45	1.0	NaN	1.0	1.0	2.0	1.0	1.0	2.0

Figura 1 - Prévia do conjunto de dados utilizado. Fonte: Elaborada pelo próprio autor (2023)

Resultados

Realizada a coleta do conjunto de dados utilizado neste trabalho, foram filtrados apenas os casos que tiveram como diagnóstico a SRAG por COVID-19, totalizando assim 3.017.888 casos. A base de dados contém informações como a semana de infecção, os sintomas, as comorbidades de pacientes, bem como da anamnese. Além disso, foram selecionados os pacientes que tiveram desfecho, ou seja, curados ou óbitos, resultando em 2.001.919 casos. Após a filtragem dos dados, percebeu-se a falta de informações principalmente em variáveis relacionadas com a presença de comorbidades como evidenciado na Tabela 2. A tabela apresenta as colunas com as maiores quantidades de valores faltantes, os quais ultrapassam mais da metade do tamanho do conjunto de dados. Essa falta de dados prejudica a análise e as conclusões baseadas nessas informações, uma vez que essas colunas poderiam ser representativas e contribuir para o processo de modelagem. Nos trabalhos de ZHOU *et al.* (2021), FERNÁNDEZ GARCÍA *et al.* (2020) e OVALLE *et al.* (2021), por exemplo, são destacados que variáveis como diabetes e obesidade estiveram presentes nos modelos e se mostraram importantes para o agravamento da doença.

Comorbidade [Nome da variável]	Valores Nulos	% Valores Nulos
Escolaridade [CS_ESCOL_N]	1.274.711	63,74%
Doença Hepática Crônica [HEPATICA]	1.272.820	63,64%
Doença Hematológica Crônica [HEMATOLOGI]	1.270.804	63,54%
Imunodeficiência [IMUNODEPRE]	1.263.254	63,16%
Asma [ASMA]	1.260.568	63,03%

Tabela 2 - Quantidade e porcentagem de valores nulos das variáveis com os maiores percentuais. Fonte: Elaborada pelo próprio autor (2023)

Contudo, mesmo com a falta de informações, optou-se por não imputar valores, ou seja, caso a variável fosse preenchida pelo valor “Ignorado” ou estivesse nula, a linha toda daquela informação seria retirada, tendo como consequência um menor número de informações na base de dados. Logo, ao adotar esse critério, o conjunto de variáveis utilizado na etapa de modelagem não pode conter valores nulos. Analisan-

do-se os dados dos infectados que possuíam desfecho da SRAG por COVID-19, verificou-se que 66,625% ($n=1.333.779$) foram curados e 33,375% ($n=668.140$) faleceram. Observando-se o sexo dos infectados, foi verificado que o sexo masculino possui uma frequência maior, sendo 55,12% ($n=1.103.476$) contra 44,86% ($n=898.211$) feminino. Também foi registrado nessa variável o valor 'Ignorado', 232 vezes. Entre os casos por gênero, observou-se que 32,8% das mulheres foram registrados como óbito, enquanto esse número foi de 33,8% no sexo masculino dos desfechos analisados. Ao examinar a faixa etária dos infectados, observou-se que, aproximadamente, 70,9% dos infectados possuem entre 40 e 80 anos e a idade média dos óbitos foi de 66 anos, enquanto a dos curados foi de 53 anos. A visualização da distribuição por idade dos curados e óbitos pode ser observada na Figura 2. Visualmente, é possível notar que a concentração dos óbitos ocorre principalmente entre os indivíduos infectados com idade mais avançada, fato esse também mencionado por ALBITAR *et al.* (2020).

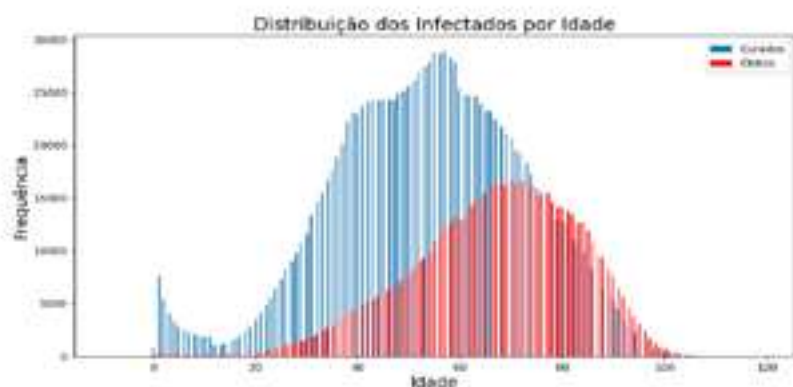


Figura 2. Distribuição dos Infectados por Idade. Fonte: Elaborada pelo próprio autor (2023)

Modelagem

Na Figura 3 é apresentada a evolução do número de pessoas infectadas pela SRAG devido à COVID-19 ao longo das semanas desde o início da pandemia em 2020. A linha azul representa o número de casos de cura, enquanto a linha verde indica o nú-

mero de óbitos registrados. É importante ressaltar que essa representação é um recorte limitado da realidade baseada nos dados disponíveis. Também é possível perceber que com o passar do ano de 2022, a quantidade de casos diminuiu significativamente, sobretudo, após março, quando 70% da população brasileira estava vacinada com a segunda dose da vacina para COVID-19, estimado no dia 01/02/2022. Para observar o impacto do cenário antes e depois da vacinação, dois modelos foram elaborados, um compreendendo o período de 01/01/2021 até 01/02/2022 e outro de 01/02/2022 até 31/12/2022.

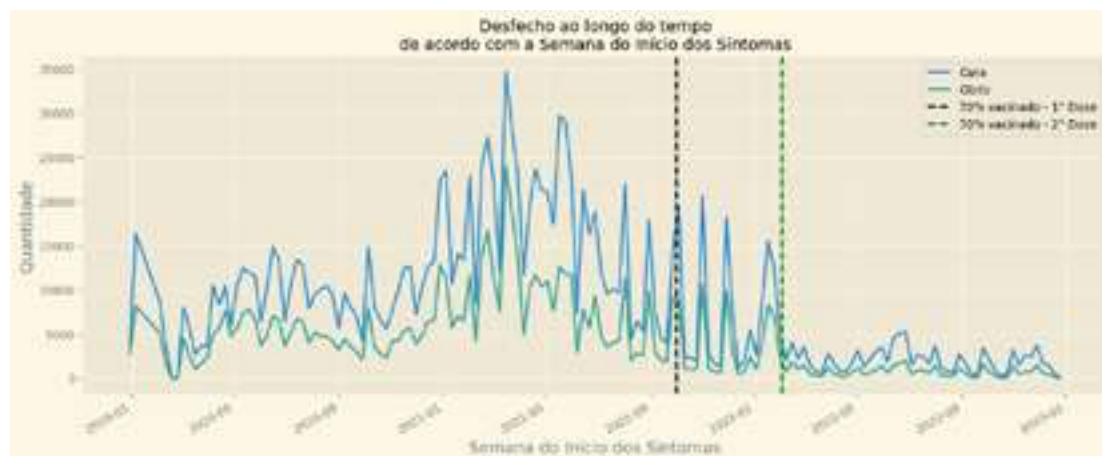


Figura 3 - Comparação entre a quantidade de curados e óbitos ao longo do tempo.

Fonte: Elaborada pelo próprio autor (2023)

Com os períodos definidos, o próximo passo foi identificar quais variáveis seriam selecionadas. Para isso, utilizou-se o algoritmo de floresta aleatória para indicar a importância de cada variável para o modelo. O algoritmo recebe como entrada um conjunto de dados contendo informações sociodemográficas e fisiológicas, totalizando 31 variáveis que foram testadas. A saída dessa etapa teve como resultado uma lista ordenada decrescentemente da importância das variáveis, segundo o algoritmo. Posteriormente, essa lista de variáveis ordenadas foi adicionada em cada modelo de forma iterativa: uma variável é adicionada a cada iteração, testada com as demais e, caso seja relevante para o modelo, é adicionada. Enquanto o algoritmo de floresta aleatória foi responsável por retornar a importância de cada variável, a regressão logística foi utilizada para de fato criar o modelo para prever a mortalidade dos infectados.

Para medir a relevância da variável em um modelo, foi verificado se a porcentagem de nulos dela não era maior que 15% da base de dados e se pelo menos duas destas três métricas: acurácia, recall ou AUCROC melhoram ao adicionar essa variável. Essas decisões foram feitas para os 2 modelos. O modelo de regressão logística que compreendia o período antes do marco de 70% de vacinados com a segunda dose da vacina para COVID-19, foi chamado de modelo 1. Esse modelo contou com 1.536.952 dados de treinamento e 61.270 de testes. Nesse modelo, utilizou-se as variáveis listadas onde seus coeficientes podem ser visualizados na Tabela 3:

- SUPPORT_VEN_INV: Se utilizou o suporte ventilatório invasivo;
- NU_IDADE_N: Idade do infectado;
- SUPPORT_VEN_NAO_INV: Se utilizou o suporte ventilatório não invasivo;
- SEM_PRI_TOTAL: Semana Epidemiológica do início dos sintomas desde o início da pandemia.

Variável	Razão de Chances	Coefficiente	IC
SUPPORT_VEN_INV	22,78	2,3203	2,308 – 2,332
SUPPORT_VEN_NAO_INV	1,69	-0,1811	-0,018
NU_IDADE_N	1,04	0,0071	0,007 – 0,007
SEM_PRI_TOTAL	0,99	-0,0255	-0,001

Tabela 3 - Razão de chance, coeficientes e intervalo de confiança para as variáveis do modelo 1. Fonte: Elaborada pelo próprio autor (2023)

Para o modelo 1 a acurácia foi de 71,8% e a área sob a curva ROC de 81,8%. Na Tabela 4 é possível verificar que o modelo alcançou uma precisão de 55% para a classe de óbitos, o que significa que acertou corretamente 55% das predições feitas para essa classe. Além disso, obteve uma revocação de 76% para a classe de óbitos, o que indica que acertou 76% das instâncias que realmente eram óbitos. A revocação é uma métrica importante para o estudo, pois mede a capacidade do modelo em identificar corretamente os casos positivos (nesse caso, os óbitos), fornecendo uma visão do quão bem o modelo está realizando essa tarefa específica.

Classe	Precisão	Revocação	Quantidade
Curados	0,85	0,7	41.100
Óbitos	0,55	0,76	20.170

Tabela 4 - Métricas do modelo do período antes do marco de 70% dos vacinados com a segunda dose da vacina para COVID-19. Fonte: Elaborada pelo próprio autor (2023.)

Para o modelo 2 também foi utilizado a regressão logística, esse modelo compreendia o período depois do marco de 70% dos vacinados com a segunda dose da vacina para COVID-19, dispondo de 102.710 dados na etapa de treinamento e 7.933 na etapa de teste. É possível destacar que a quantidade de dados presentes para treinar o modelo diminuiu significativamente devido ao corte no período dos dados. Quanto aos resultados, o modelo 2 teve as mesmas variáveis selecionadas do modelo anterior no método de seleção passo a passo para frente, mas com coeficientes do modelo diferentes, conforme apresentado na Tabela 5. Ao examinar as métricas no conjunto de teste, observou-se uma melhor acurácia (80%) e valores maiores nas demais métricas para a classe dos curados, enquanto o modelo 1 apresentou valores para as melhores métricas para a classe dos óbitos. Para completar, a métrica AUCROC foi de 80,9%, valor próximo do modelo 1. As métricas individuais de cada classe podem ser encontradas na Tabela 6.

Variável	Razão de Chances	Coefficiente	Intervalo de Confiança
SUPPORT_VEN_INV	30,29	3,2695	3,217 - 3,322
SUPPORT_VEN_NAO_INV	2,53	0,8441	0,804 - 0,884
NU_IDADE	1,03	0,0284	0,028 - 0,029
SEM_PRI_TOTAL	0,98	-0,0313	-0,032 - -0,031

Tabela 5 - Coeficientes, razão de chances e o intervalo de confiança para as variáveis do modelo 2. Fonte: Elaborada pelo próprio autor (2023)

Classe	Precisão	Revocação	Quantidade
Curados	0,9	0,74	3.611
Óbitos	0,45	0,71	1.068

Tabela 6 - Métricas do modelo do período depois do marco de 70% dos vacinados com a segunda dose da vacina para COVID-19. Fonte: Elaborada pelo próprio autor (2023)

O destaque de ambos modelos foi para as variáveis relacionadas ao uso do suporte ventilatório. O suporte ventilatório invasivo teve o valor de 22,78 para a razão de chances no modelo 1 e 30,29 para o modelo 2. Ou seja, caso o paciente precise utilizar esse suporte ventilatório, sua chance de óbito aumentaria consideravelmente mesmo após o período com 70% dos vacinados, evidenciando-se uma forte importância dessa variável neste trabalho. Fato esse destacado também no trabalho de MARINI & GATTINONI (2020), de acordo com o autor, o uso do suporte ventilatório pode piorar a condição de pacientes com COVID-19.

Vale ressaltar que essa piora pode ser por causa da própria situação do paciente, visto que o mesmo já se encontra em situação agravada por diversos fatores e necessita do procedimento mecânico para ajudar na respiração para sobreviver. Um dado interessante a se destacar é que 75% dos infectados após o marco de 70% dos vacinados com a segunda dose para COVID-19, que precisaram utilizar o suporte ventilatório invasivo morreram, reforçando a ideia de que não necessariamente seja o suporte ventilatório invasivo o principal causador das mortes, mas provavelmente, um estado já agravado por outros fatores anteriores e, que ao tentar utilizar o suporte ventilatório invasivo com esse quadro clínico, o paciente pode vir a desenvolver outras doenças como por exemplo a pneumonia e não resistir.

Outro destaque é para a variável de idade, com coeficientes de 1,04 e 1,03 para os modelos 1 e 2 respectivamente. Essa informação mostra que, a cada ano mais velho, a pessoa teria cerca de 4% mais chance de morrer antes do marco de 70% de vacinação para a segunda dose. Uma observação interessante é que, mesmo após essa data de referência da vacinação, o indivíduo infectado ainda teria um aumento de 3% na probabilidade de óbito. Essa variável também foi importante e contribuiu para a mortalidade nos casos da COVID nos trabalhos de LIU *et al.* (2020) e ALBITAR *et al.* (2020). Embora esses trabalhos sejam referentes a COVID e não especifiquem a análise da SRAG causada pela COVID-19, é possível identificar a interseção de ideias com o trabalho atual. Logo, destaca-se o cuidado com as pessoas mais idosas no enfrentamento da SRAG pela COVID-19 a fim de evitar o agravamento da doença. Além disso, no modelo 2 houve uma redução da razão de chances para cada ano na idade do paciente. Isto sugere uma importante mudança após o marco de vacinação, especialmente para as pessoas mais velhas. Em relação às métricas utilizadas para avaliar os

modelos, verificou-se que a métrica AUCROC presente na literatura entre os trabalhos relacionados para a modelagem de dados da COVID-19 permeia em torno de 0,73 a 0,84 (BARDA *et al.*, 2020). No presente trabalho a curva AUROC foi de 0,81 estando dentro da faixa de desempenho apresentada na literatura. Analisando-se a acurácia, percebe-se que alguns trabalhos relacionados tiveram métricas entre 72% e 82% (HU *et al.*, 2021; FARIA, 2021; WOLLENSTEIN-BETECH *et al.*, 2020), contra 80% desse trabalho. Entretanto, vale ressaltar dois pontos no trabalho de HU *et al.* (2021) que teve uma acurácia de 82%. Em primeiro lugar, o estudo utilizou variáveis laboratoriais que não estão presentes na base de dados utilizada. A inclusão dessas variáveis poderia melhorar os resultados, uma vez que fornecem informações adicionais sobre o paciente. Como segundo ponto, destaca-se o tamanho da base de dados utilizada para construir o modelo. HU *et al.* (2021) utilizou 422 dados em seu modelo, enquanto para esse trabalho, o modelo 2 utilizou-se 102.710 dados para treinamento, que, consequentemente, proporciona maior robustez aos modelos aqui apresentados.

Resultados

Conforme exposto, o presente estudo possibilitou a comparação entre os períodos pré-vacinação e pós-vacinação, bem como a identificação das características mais relevantes que impactaram na classificação da mortalidade em indivíduos afetados pela SRAG ocasionada pela COVID-19. Desse modo, mesmo diante de uma ampla quantidade de dados ao longo do tempo e diversas variáveis, foi possível determinar o período e as variáveis mais importantes para construir os modelos que sejam mais próximos da situação atual, permitindo, consequentemente, a tomada de decisão mais rápida e efetiva. Uma contribuição significativa deste trabalho foi evidenciar o comportamento das variáveis selecionadas em relação ao marco de vacinação. Isto resultou em uma redução do impacto da idade e aumentou a importância do suporte ventilatório como fator determinante para o desfecho de óbito. Como limitação, o presente estudo aponta a ausência de dados clínicos e laboratoriais, refletindo na impossibilidade de fazer inferências acerca dos pacientes por meio de marcadores biológicos. Além disso, há um número considerável de dados faltantes nos formulários preenchidos, o que limita as análises. Apesar das limitações mencionadas, o presente estudo apresenta um tama-

nho amostral robusto, o que possibilita inferências confiáveis sobre os dados coletados e contribui para a tomada de decisões na conduta quanto aos pacientes. Como perspectivas para futuras pesquisas, pode-se citar o melhoramento da coleta de dados de forma a reduzir a fragilidade da base de dados utilizada e investigação mais aprofundada da relevância do suporte ventilatório invasivo no contexto da Síndrome Respiratória Aguda Grave causada pela COVID-19 nos casos de óbito.

Considerações finais

Este trabalho foi parcialmente financiado pela Coordenadoria de Aperfeiçoamento do Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico, Tecnológico (CNPq) e Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), e Instituto federal do Norte de Minas Gerais (IFNMG).

Referências

AGUIAR, P.; NUNES, B. Odds Ratio: Reflexão sobre a Validade de uma Medida de Referência em Epidemiologia. *Acta medica portuguesa*, v. 26, n. 5, p. 505–510, 2013.

ALBERT, S.; LINVILLE, L. Benchmarking current and emerging approaches to infrasound signal classification. *Seismological research letters*, v. 91, n. 2A, p. 921–929, 2020.

ALBITAR, O. et al. Risk factors for mortality among COVID-19 patients. *Diabetes research and clinical practice*, v. 166, n. 108293, p. 108293, 2020.

BARDA, N. et al. *Performing risk stratification for COVID-19 when individual level data is not available – the experience of a large healthcare organization*. 2020. Disponível em: <<http://dx.doi.org/10.1101/2020.04.23.20076976>>.

BREIMAN, L. *Machine learning*, v. 45, n. 1, p. 5–32, 2001.

CARVALHO, J. A. M. DE; RODRÍGUEZ-WONG, L. L. *A transição da estrutura etária da população brasileira na primeira metade do século XXI*. *Cadernos de saúde pública*, v. 24, n. 3, p. 597–605, 2008.

CHICCO, D.; JURMAN, G. *Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone*. *BMC medical informatics and decision making*, v. 20, n. 1, p. 16, 2020.

CORONAVIRIDAE STUDY GROUP OF THE INTERNATIONAL COMMITTEE ON TAXONOMY OF VIRUSES et al. *The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2*. *Nature microbiology*, v. 5, n. 4, p. 536–544, 2020.

CUCINOTTA, D.; VANELLI, M. *WHO declares COVID-19 a pandemic*. *Acta bio-medica : Atenei Parmensis*, v. 91, n. 1, p. 157–160, 2020.

DELONG, E. R.; DELONG, D. M.; CLARKE-PEARSON, D. L. *Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach*. *Biometrics*, v. 44, n. 3, p. 837–845, 1988.

DIEZ, D.; ÇETINKAYA-RUNDEL, M.; BARR, C. *OpenIntro statistics*. [s.l.: s.n.].

DREISEITL, S.; OHNO-MACHADO, L. *Logistic regression and artificial neural network classification models: a methodology review*. *Journal of biomedical informatics*, v. 35, n. 5–6, p. 352–359, 2002.

DUDA, R. O.; HART, P. E.; STORK, D. G. *Pattern Classification*. 3. ed. [s.l.] Standards Information Network, 2022.

Elkan, C.P. (1997). *Boosting and Naive Bayesian learning*.

FARIA, A. R. Q. DE P. *Análise de sobrevivência e fatores prognósticos associados à mortalidade em pa-*

cientes com SRAG por Covid-19 hospitalizados em UTI na Paraíba. 2021.

FERNÁNDEZ, A. et al. *Learning from imbalanced data sets*. 1. ed. Basel, Switzerland: Springer International Publishing, 2018.

FERNÁNDEZ GARCÍA, L.; PUENTES GUTIÉRREZ, A. B.; GARCÍA BASCONES, M. *Relationship between obesity, diabetes and ICU admission in COVID-19 patients. Medicina Clínica (English Edition)*, v. 155, n. 7, p. 314–315, 2020.

FRANCESCHI, P. R. DE. *Modelagens preditivas de Churn: o caso do Banco do Brasil*. 2019.

GORBALENYA, A. E. et al. *Severe acute respiratory syndrome-related coronavirus: The species and its viruses – a statement of the Coronavirus Study Group*. 2020. Disponível em: <<http://dx.doi.org/10.1101/2020.02.07.937862>>.

GUDE-SAMPEDRO, F. et al. *Development and validation of a prognostic model based on comorbidities to predict COVID-19 severity: a population-based study*. *International journal of epidemiology*, v. 50, n. 1, p. 64–74, 2021.

HASTIE, T.; TIBSHIRANI, R.; TIBSHIRANI, R. J. *Extended comparisons of best subset selection, forward stepwise selection, and the lasso*. 2017. Disponível em: <<http://arxiv.org/abs/1707.08692>>.

HU, J.; FEI, Y.; LI, W.-Q. *Predicting the mortality risk of acute respiratory distress syndrome: radial basis function artificial neural network model versus logistic regression model*. *Journal of clinical monitoring and computing*, v. 36, n. 3, p. 839–848, 2022.

HUANG, C. et al. *Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China*. *Lancet*, v. 395, n. 10223, p. 497–506, 2020.

JAMES, G.; WITTEN, D.; HASTIE, T. *An introduction to statistical learning: With applications in R*. New York, NY: Springer, 2013.

KÜCHEMANN, B. A. *Envelhecimento populacional, cuidado e cidadania: velhos dilemas e novos desafios*. *Sociedade e Estado*, v. 27, n. 1, p. 165–180, 2012.

LI, X. et al. *Molecular immune pathogenesis and diagnosis of COVID-19*. *Journal of pharmaceutical analysis*, v. 10, n. 2, p. 102–108, 2020.

LIMA, T. P. F. et al. *Death risk and the importance of clinical features in elderly people with COVID-19 using the Random Forest Algorithm*. *Revista Brasileira de Saúde Materno Infantil*, v. 21, n. suppl 2, p. 445–451, 2021.

LIU, W. et al. *Analysis of factors associated with disease outcomes in hospitalized patients with 2019 novel coronavirus disease*. *Chinese medical journal*, v. 133, n. 9, p. 1032–1038, 2020.

LU, R. et al. *Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding*. *Lancet*, v. 395, n. 10224, p. 565–574, 2020.

LUO, H. et al. *Logistic regression and random forest for effective imbalanced classification*. 2019 *IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC)*. *Anais...IEEE*, 2019.

M., C. *Pattern Recognition and Machine Learning*. Nova Iorque, NY, USA: Springer, 2016.

MAIMON, O.; ROKACH, L. (EDS.). *Data mining and knowledge discovery handbook*. 2. ed. New York, NY: Springer, 2010.

MARINI, J. J.; GATTINONI, L. *Management of COVID-19 respiratory distress*. *JAMA: the journal of the American Medical Association*, v. 323, n. 22, p. 2329, 2020.

MARSONO, M. N.; EL-KHARASHI, M. W.; GEBALI, F. *Targeting spam control on middleboxes: Spam detection based on layer-3 e-mail content classification*. *Computer networks*, v. 53, n. 6, p. 835–848, 2009.

MONTAZERI, M. et al. *Machine learning models in breast cancer survival prediction*. *Technology and health care: official journal of the European Society for Engineering and Medicine*, v. 24, n. 1, p. 31–42, 2016.

MONTGOMERY, D. C.; PECK, E. A. *Introduction to Linear Regression Analysis*. 3. ed. Nashville, TN: John Wiley & Sons, 2001.

MUSTAFA ABDULLAH, D.; MOHSIN ABDULAZEEZ, A. *Machine learning applications based on SVM classification A review*. *Qubahan Academic Journal*, v. 1, n. 2, p. 81–90, 2021.

OVALLE, D. L. P. et al. COVID obesity: A one-year narrative review. *Nutrients*, v. 13, n. 6, p. 2060, 2021.

PATIL, K. et al. Deep learning based car damage classification. 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA). *Anais...IEEE*, 2017.

PRATI, R. C.; BATISTA, G. E. A. P. A.; MONARD, M. C. Evaluating classifiers using ROC curves. *IEEE Latin America Transactions*, v. 6, n. 2, p. 215–222, 2008.

PROVOST, R. K. Glossary of terms. *Journal of Machine Learning*, v. 30, p. 271–274, 1998.

QUINLAN, J. R. Bagging, Boosting, and C4. 5. *Proceedings of the AAAI Conference on Artificial Intelligence*, v. 13, 1996.

RÄTSCH, G. A brief introduction into Machine Learning. 2004.

RIOS, L. F. B. Modelos de predição de risco de morte para pacientes com carcinoma epidermoide de cabeça e pescoço. [s.l.] Universidade de Sao Paulo, Agencia USP de Gestao da Informacao Academica (AGUIA), 2021.

RUSSELL, S. J.; NORVIG, P. Artificial intelligence: A modern approach. [s.l.] Prentice Hall, 2010.

SENA, G. R. (2021). Modelos Preditivos de Óbito para Pacientes de Óbito para Pacientes com COVID-19.

SHARMA, S. *Applied Multivariate Techniques*. Nashville, TN: John Wiley & Sons, 1995.

TANBOBA, I. H. et al. Development and validation of clinical prediction model to estimate the probability of death in hospitalized patients with COVID-19: Insights from a nationwide database. *Journal of medical virology*, v. 93, n. 5, p. 3015–3022, 2021.

WANG, H.; MA, C.; ZHOU, L. A brief review of machine learning and its application. 2009 International Conference on Information Engineering and Computer Science. *Anais...IEEE*, 2009.

WOLLENSTEIN-BETECH, S. et al. Physiological and socioeconomic characteristics predict COVID-19

mortality and resource utilization in Brazil. PloS one, v. 15, n. 10, p. e0240346, 2020.

ZHOU, Y. et al. Obesity and diabetes as high-risk factors for severe coronavirus disease 2019 (Covid-19).

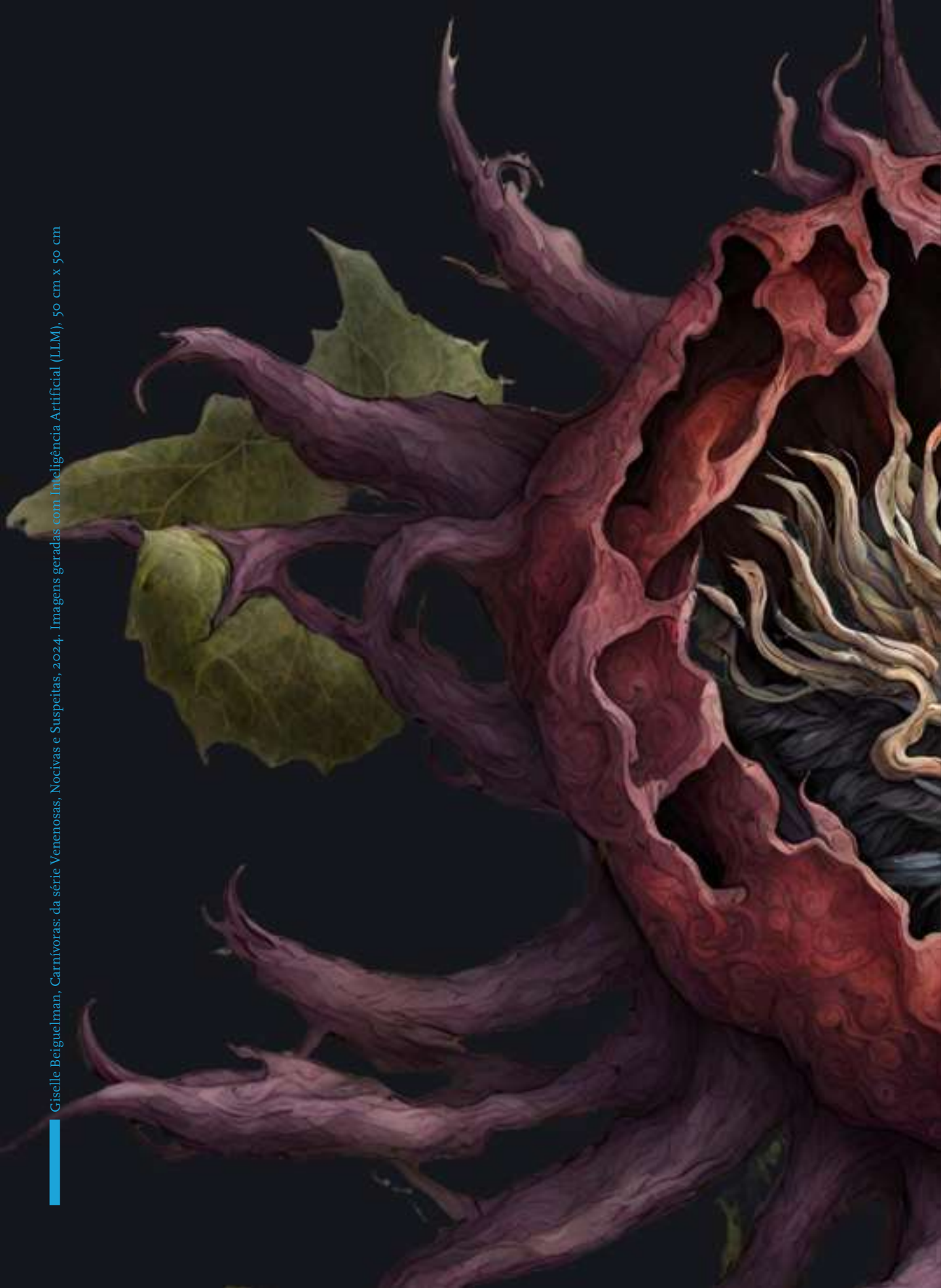
Diabetes/metabolism research and reviews, v. 37, n. 2, p. e3377, 2021.



■ Giselle Beiguelman, frames do vídeo generativo da série Flora Rebellis, versão 2024.
Imagens geradas com Inteligência Artificial (Style Gans)

■ [Acesse o vídeo clicando na imagem](#)





REFLEXÕES EPISTEMOLÓGICAS E ÉTICAS SOBRE A INTELIGÊNCIA ARTIFICIAL COMO ÁREA DA COMPUTAÇÃO COGNITIVA

DOI: <https://doi.org/10.35699/2965-6931.2023.47723>

MARIAH BROCHADO*; LUCAS MAGNO DE OLIVEIRA PORTO**

RESUMO: O texto ora apresentado suscita algumas considerações sobre o progresso científico que possibilitou o surgimento de um campo da Ciência da Computação, a chamada computação cognitiva, cujo feito maior é o advento da Inteligência Artificial, em torno da qual pesquisas e criações tecnológicas vêm mudando radicalmente o ethos humano em proporções jamais experimentadas. Além da abordagem relativa à cientificidade que molda a concepção de mente como metabolismo do cérebro, pretende-se ainda apontar a necessidade de se refletir sobre tais ocorrências no âmbito da Ética Filosófica (Filosofia Prática), eis que parece no mínimo perigoso reduzir epistemologicamente a integralidade da condição humana a mecanismos biológicos. Esta tendência vem sendo atualizada nas discussões sobre IA e agenciamento maquínico, tendência este que subtrai do humano sua espiritualidade ao equipará-lo a entidades maquinicas que simulam sedutoramente os mecanismos do cérebro do sapiens.

PALAVRAS-CHAVE: cientificidade; modernidade; ética; computação cognitiva; inteligência artificial.

Epistemological and ethical reflections on artificial intelligence as an area of cognitive computing

ABSTRACT: This paper raises some considerations about the scientific progress that enabled the development of a Computer Science's research field, the so-called cognitive computing, whose greatest achievement is the advent of Artificial Intelligence, around which research and technological creations have been radically changing the human ethos in unprecedented proportions. In addition to the approach relating to scientificity that shapes the conception of the mind as metabolism of the brain, it is also intended in this study to seek the need to reflect on such occurrences in the scope of Philosophical Ethics (Practical Philosophy), since it seems to at least reduce epistemologically the integrality of the human condition to biological manipulation. This trend has been updated in discussions about AI and machinic agency, a trend that deprives humans of their spirituality by equating them to machinic entities that seductively simulate the mechanisms of the sapiens brain.

KEYWORDS: scientificity; modernity; ethics; cognitive computing; artificial intelligence.

* Universidade Federal de Minas Gerais / ** Universidade Federal de Minas Gerais

Introdução

A experiência técnica contemporânea impõe à Filosofia de modo geral, e à Ética em particular, reflexões sobre temas jamais cogitados, seja no âmbito das ciências, seja no âmbito da tecnologia, porque jamais postos nos termos em que hoje se anunciam: entre eles, o mais destacado e preocupante, prova e feito maior da evolução técnica até nossos dias, é a noção de inteligência artificial.

Trata-se, grosso modo, de uma simulação maquínica em computadores hiperpotentes, a partir de dados acumulados por décadas, para reproduzir elementos constitutivos da racionalidade humana, o que se faz acompanhar da expectativa de que máquinas de alta performance cognitiva possam vir a se autonomizar independentemente do que chamamos intencionalidade, característica aparentemente exclusiva – ao menos até este momento na escala evolutiva dos mamíferos – dos humanos. Essa expectativa talvez seja a sanha mais persuasiva do discurso futuroológico de cientistas da computação e de programadores hoje, retórica contagiante a prospectar a revolução que será promovida no campo das Neurociências, ao se associarem à engenharia reversa do cérebro, para seguir rumo à incorporação humanoide da chamada inteligência artificial forte (John Searle), possibilitada pelo poder computacional aplicado ao campo da Robótica.

Nesse contexto, importa refletir sobre as bases a partir das quais se formou a área da Computação Cognitiva, cujo feito mais exuberante é justamente a arquitetura da área nomeada hoje Inteligência Artificial, a qual no passado se chamava singelamente Simulação Computacional, da mesma forma que um dia o que hoje nomeamos Computador foi denominado Cérebro Eletrônico. A ideia de que o cérebro possa ser reproduzido artificialmente e até mesmo emulado em novas formas de inteligência maquínica é uma nova inspiração para a releitura da tradicional relação homem-máquina, a qual há muito é relatada na história da filosofia, particularmente na figura de autômato, desde Aristóteles.

Ocorre que neste século a filosofia tradicional de base racionalista perdeu muito do seu espaço na análise de fenômenos maquínicos sem precedentes em nossa historiografia, impactando a própria concepção de humano, hoje dissecado como objeto submetido à verificabilidade científica como qualquer fenômeno causal. Em se tratando

do de cognição humana, é o cérebro que ganha protagonismo nesse processo, vindo a substituir a insondável natureza humana, pressuposto da filosofia tradicional.

Fato é que o surgimento de uma nova área da computação cognitiva, qual seja, a Inteligência Artificial, vem mudando radicalmente o ethos humano, razão por que, além da abordagem relativa à cientificidade que molda teorias da mente e do cérebro, pretende-se ainda apontar a necessidade de se refletir sobre tais ocorrências no âmbito da Filosofia Prática (ou Ética), sob pena de reduzirmos a integralidade humana a mecanismos biológicos, negando à mesma qualquer espiritualidade, no que a igualaríamos a entidades maquinicas que simulam artificialmente os mecanismos da cognição humana.

O cenário epistemológico passado e presente na distinção entre cérebro e mente

A disputa entre a versão metafísica de mente como irreduzível ao cérebro e as novas tendências da neurofisiologia e da computação cognitiva em reduzir os estados mentais ao metabolismo cerebral remonta ao dualismo cartesiano e à inferência do estado de consciência como um dado imediato da experiência humana que independe de demonstração – pela evidência do “cogito, ergo sum” cartesiano, vale dizer, a experiência do pensar nos leva à inferência de que existimos. Como propõe Descartes (1657, p. 17-32), posso duvidar de qualquer item pensado, mas não posso duvidar de que estou pensando. Exemplifica Gerard Lebrun em comentário acerca das reflexões do Filósofo em A Segunda Meditação:

Quando percebo o pedaço de cera, seja compreendendo clara e distintamente sua natureza, seja apenas imaginando-o ou tocando-o, só uma coisa é certa no ponto em que me encontro. É que eu penso percebê-lo... Mostrando que este ‘pensamento’ era indispensável ao conhecimento da coisa, a análise precedente deu confirmação a esta verdade. (Descartes, 1983, p. 106).

A busca pela base orgânica da faculdade de pensar não é novidade na história da

filosofia, inclusive o próprio Descartes teria reconhecido a existência de um conector (glândula pineal) entre corpo e espírito, segundo o qual a espiritualidade humana, independente do corpo, poderia influenciá-lo. Como referenciado em John Cottingham, Descartes assim se manifesta em *As paixões*:

[...] é preciso reconhecer que, embora a alma esteja unida ao corpo como um todo, há neste uma parte em que ela exerce as suas funções mais particularmente do que em todas as outras [...] Essa parte não é o coração, ou o cérebro no seu conjunto, mas apenas a parte mais interior do cérebro, que é uma pequena glândula, situada no meio de sua substância e acima do canal por onde os espíritos das suas concavidades anteriores comunicam-se com aqueles das suas concavidades posteriores” [...] A consciência sensorial tem lugar quando a alma “inspeciona” uma imagem que está literalmente impressa na glândula. (Cottingham, 1995, p. 74).

Kant também se ocupou, ainda que nos estritos limites do seu criticismo transcendental, do problema de a formação do conhecimento situar-se numa instância mediadora entre a experiência e o conceito, o que está registrado em um dos capítulos mais obscuros de *A Crítica da Razão Pura*, o qual foi reservado ao Esquematismo Transcendental. Nele, Kant esboça a noção de imagem e modelo, talvez a versão mais aproximada das formulações de uma Filosofia da Mente, evidentemente sem a conotação cientificista empírica que influencia esta, como veremos. De qualquer modo, trata-se de uma referência relevante na tradição filosófica racionalista, haja vista que, no bojo da sua filosofia transcendental, Kant admite uma instância em que imagens são formuladas, o que permite a conexão entre intuições sensíveis e categorias do intelecto, instância esta que ele denomina imaginação transcendental. No entanto, evidentemente por não ter nenhuma base científica para discutir a estrutura do cérebro e a formação dos estados mentais (à sua época só existia a psicologia racional, cujo suposto é lógico, e não observacional), o filósofo postulou a existência de uma instância conectiva entre experiência sensível e inteligível, em que pese não ser propósito do seu criticismo esse tipo de abordagem, o que era impensável à época, razão por que Kant considera essa instância insondável, inexplicável, sendo a temática pouco explorada em sua obra. O filósofo dedica a ele poucas páginas na monumental estrutura de *A Crítica da Razão Pura*.

Em apertada síntese, registramos que o tema é apresentado por Kant como uma capacidade que o ser humano tem de montar esquemas gerais de percepção dos objetos reais a partir da imagem que formula deles. Ocorre que o esquema mental não se confunde com as diversas imagens que o inspiraram e sobre as quais ele retorna. Por exemplo, uma imagem de cinco pontos alinhados me leva a uma imagem do número cinco, mas, quando penso em um número qualquer, este não se vincula a qualquer imagem. Eis a razão por que sou capaz de pensar o número 1000, mas é impossível que eu consiga abranger a imagem de mil pontos alinhados. A própria imagem de 1000 é uma representação do processo geral da imaginação para dar um conceito a essa imagem, independentemente da quantidade de pontos que possamos imaginar. Essa imagem que representa um conceito e que servirá para se adequar a qualquer outra imagem que ela representa é o que Kant chama de esquema, objeto de suas reflexões sobre o esquematismo transcendental. Ele distingue, portanto, imagem e esquema (Kant, 2014, p. 224).

Raramente Kant exemplifica quando está expondo seus argumentos, mas ele traz um exemplo para melhor elucidar essa capacidade da faculdade intelectual humana de formular imagens e esquemas gerais a partir delas, esquemas estes que não se restringem a nenhuma das imagens às quais se aplicam. Para Kant esse fenômeno da inteligência humano é inexplicável, “uma arte oculta nas profundezas da alma humana, cujo segredo de funcionamento dificilmente poderemos alguma vez arrancar à natureza e pôr a descoberto perante os nossos olhos”. Eis o exemplo para elucidar essa capacidade esquematizadora atribuída por ele à “alma” humana:

O conceito de cão significa uma regra segundo a qual a minha imaginação pode traçar de maneira geral a figura de certo animal quadrúpede, sem ficar restringida a uma única figura particular que a experiência me oferece ou também a qualquer imagem possível que posso representar in concreto [...] Só poderemos dizer que a imagem é um produto da faculdade empírica da imaginação produtiva, e que o esquema de conceitos sensíveis (como das figuras no espaço) é um produto e, de certo modo, um monograma da imaginação pura a priori, pelo qual e segundo o qual são possíveis as imagens; estas, porém, têm de estar sempre ligadas aos conceitos, unicamente por intermédio do esquema que elas designam e ao qual não são em si mesmas inteiramente adequadas. (Kant, 2010: 183/B180-A141).¹

¹ No original: “Der Begriff vom Hunde bedeutet eine Regel, nach welcher meine Einbildungskraft die Gestalt eines vierfüßigen Tieres allgemein verzeichnen kann, ohne auf irgend eine einzige besondere Gestalt, die mir die Erfahrung darbietet, oder auch ein jedes mögliches Bild, was ich in concreto darstellen kann, eingeschränkt zu sein. [...] So viel können wir nur sagen: das Bild ist ein Produkt des empirischen Vermögens der produktiven Einbildungskraft, das Schema sinnlicher Begriffe (als der Figuren im Raume) ein Produkt und gleichsam ein Monogramm der reinen Einbildungskraft a priori, wodurch und wornach die Bilder allererst möglich werden, die aber mit dem Begriffe nur immer vermittelt des Schema, welches sie bezeichnen, verknüpft werden müssen, und an sich demselben nicht völlig kongruieren. (KANT, 2014, p. 224-225, grifo do autor).”

Posteriormente a Kant, já no final do século XIX, desenvolvem-se teorias biológicas do conhecimento, tal como a proposta por Richard Ludwig Avenarius e Ernst Mach. Estes apresentaram uma leitura biológica do processo de conhecimento, sendo este uma função vital como qualquer outra. Encontramos também teses que defendiam a lei da heterogenia dos fins (Vaihinger) e que chegaram a se ocupar da hipótese de redução da racionalidade humana a processos biológicos e filogenéticos (Brochado, 2021; Mora, 2001, p. 2965).

Insta lembrar que é no século XIX que floresce o positivismo, o qual evoluiu de obras inicialmente sobre mecânica (como a de Ernst Mach) para o interesse em lógica simbólica, duas tendências que, combinadas, levaram a um novo movimento em torno de professores da Universidade de Viena (como Mach e Moritz Schlick), cuja filosofia ficou conhecida como positivismo lógico, um tipo específico de movimento positivista conduzido por um grupo chamado Círculo de Viena. Segundo essa tradição,

[...] a soma total do nosso conhecimento é propiciada pela ciência [sendo] a metafísica ao estilo antigo estritamente verbosidade vazia. Não há nada que possamos conhecer além da experiência. Nisso encontramos uma certa afinidade com as ideias kantianas, se omitirmos os números. A insistência na observação empírica vem acompanhada de um critério de significado um tanto vinculado ao pragmatismo rotineiro do cientista de laboratório. Trata-se do famoso princípio da verificabilidade, ou verificacionismo, segundo o qual o significado de uma proposição é o seu método de verificação. Deriva-se de Mach, que utilizou esse tipo de procedimento ao definir os termos empregados em mecânica. (RUSSELL, 2017, p. 399).

Esta concepção que exige o método verificabilista no labor científico é a base sobre a qual será edificada a Filosofia da Mente, o que remonta à aparentemente irreconciliável distinção entre a tradição racionalista (ou continental) e a tradição empirista (ou analítica) (Russell, 2017, p. 392; D'Agostini, 1999, p. 90). Em Bertrand Russell, encontramos sucintamente uma distinção tradicional na história da filosofia ocidental: o racionalismo e o empirismo, que se distinguem, inclusive, em termos de influência territorial, uma vez que a filosofia europeia continental e a filosofia britânica seguiram caminhos radicalmente opostos no pensar filosófico, elegendo diferentes pontos de partida para pensar a realidade, chegando a pontos de vista inconciliáveis, particularmente no quesito ontologia humana.

A origem dessa cisão, de acordo com Russell, já se encontra na obra de Descartes, a partir da qual se formam duas correntes de pensamento: a do empirismo britânico, representado por Locke, Berkeley e Hume, e o antigo racionalismo revivido por Descartes, cujos legatários mais expressivos, no século XVII, foram Leibniz (na Alemanha) e Spinoza (na Holanda). Na obra *Ensaio*, John Locke tenta estabelecer, pela primeira vez, quais os limites do conhecimento humano e qual investigação está a nosso alcance, contra a convicção otimista e isenta de crítica dos racionalistas, que postulavam um conhecimento perfeito a nós acessível. A filosofia proposta por Locke é empírica em dois sentidos, base de uma filosofia crítica: primeiro, porque não prejudica o alcance do conhecimento humano, o que foi empreendido pelos racionalistas; segundo, porque dá ênfase à experiência sensorial, baseando o conhecimento estritamente na experiência, quer dizer, rejeitando as ideias inatas de Descartes e Leibniz. Conforme as palavras de Russell,

É admitido por todos que desde o nascimento possuímos alguma espécie de bagagem inata capaz de se desenvolver e que nos possibilita aprender um certo número de coisas. Mas é inútil supor que a mente inculta possui um conteúdo adormecido. Se assim fosse, jamais poderíamos distinguir entre esse e outro conhecimento genuinamente originado da experiência. Então poderíamos dizer que todo o conhecimento é inato. Isso é precisamente o que se diz na teoria da anamnese mencionada no *Ménon*. Então, para começar, a mente é como uma folha de papel em branco. O que a preenche de conteúdo mental é a experiência. (RUSSELL, 2017, p. 281).

Percebe-se uma tendência dos filósofos da mente em sustentar suas teses em fontes da tradição analítica de base empírica, seguindo a divergência apontada por Russell, no que diverge da tradição racionalista, a qual supõe uma natureza humana indemonstrável empiricamente. A questão é que a tradição empirista depende do êxito científico para em torno dele fazer gravitar sua crítica filosófica, o que no momento não soma em nada à filosofia da mente, haja vista que a relação mente/cérebro como pano de fundo para se discutir inteligência artificial é tão empiricamente frágil quanto o postulado racionalista de uma natureza humana incólume a causalismos naturais, ou o inatismo cartesiano, já que o cérebro ainda está longe de ser totalmente conhecido pelas ciências empíricas. A adoção da tese de uma natureza humana ou de um funcionamento suposto do cérebro ainda não acessado pela ciência é, *mutatis mutandis*, comprovação alguma: ambas são pressuposições que visam a justificar a existência de uma instân-

cia, sem serem capazes de demonstrar essa existência. Derrotar a concepção clássica de natureza humana sem ter onde se fiar em termos de explicação empírica do funcionamento do cérebro que gera as atividades da mente não parece ser atitude científica plausível.

Corroborando a afirmação de Russell, percebe-se uma tendência bastante evidente neste início de século quanto à convicção de que a filosofia deva ser submetida ao crivo da ciência, entendendo-se esta como conhecimento de base empírica, gerando certa ideologização das práticas científicas, denotando uma espécie de “desencantamento” com a suposta descoberta da razão na modernidade. Nas palavras de Russell: “à medida que a teoria científica tenta abranger o mundo inteiro, persegue um objetivo semelhante ao da metafísica. Onde a ciência difere é na sua maior responsabilidade para com fatos complexos e recalcitrantes” (Russell, 2017, p. 388). Substituir as explicações metafísicas da realidade, dentre elas a referência a uma natureza humana insondável, pela pretensa precisão científica de base observacional não parece ter correspondido à ambição dos modernos. Nesse ponto, importa trazer algumas reflexões trazidas por Ulrich Beck a propósito do referido desencantamento científico.

Segundo ele (Beck, 2011 p. 231, 235), uma lenda inventada no século XIX definiu a sociedade industrial, em seu esquematismo de vida e trabalho, como sociedade moderna, o que esconde o fato de que a modernidade não cumpriu seu telos de ser fonte de solução para as questões humanas. Perigos que eram antes atribuídos às forças externas e superiores à humanidade, seja na forma de eventos da natureza, seja na forma de eventos divinos, passam, com o advento da explicação científica dos fenômenos, a ser referidos à própria racionalidade humana, a qual tem a tarefa de justificar todas as ocorrências, em todas as instâncias da experiência humana, em termos de verdade universal alcançável diretamente pela razão (universal concreto, portanto, não mais abstratificado em mitos).

Se antes o peso existencial do ser humano era amenizado pela referência a um destino exterior a ele, natural ou divino, na forma de mitos e religiões (Mayos, 2021, p. 691), o processo de cientificização promete substituir tais referências e dar conta de enfrentar todos os perigos que surjam da situação natural e social do ser humano, o que acaba tornando a ciência vítima da sua própria criticidade. Isso porque, ao se dispor a explicar a natureza, o homem e a sociedade como dados preexistentes a ela,

a ciência passa de uma fase em que simplesmente enfrenta tais fenômenos para uma segunda fase, na qual se vê obrigada, por coerência com sua essência e seus princípios, a fazer incidir sua pretensão de racionalidade a si mesma, causando um desencantamento quanto às pretensões de verdade e de esclarecimento na medida em que esse exercício autorracionalizado não encontra resposta última no plano do entendimento (em Hegel será atingido no plano da razão, fora do horizonte limitado das ciências).

A primeira fase do trajeto da ciência é a que Beck (2011, p. 235) denomina fase simples, na qual a ciência só se realizou pela metade; e, na segunda fase, que ele chama de fase reflexiva, quando ocorre uma cientificização completa, o emprego metódico da dúvida científica sobre “seus próprios produtos, carências e tribulações, depara[-]se assim com uma segunda gênese civilizatória”, cuja expansão pressupõe a extensão das bases críticas da própria ciência aos seus próprios especialistas.

Nessa perspectiva das subjetividades que representam e encarnam o saber científico (os cientistas especialistas nas diversas áreas), acontece curiosamente uma desestabilização dos fundamentos da ciência, e sua autocompreensão (originariamente legitimada pela autocrítica dos próprios especialistas) passa a se vulnerabilizar na forma de autocrítica mediada publicamente, gerando na sociedade dessa civilização científica um “processo de desmistificação das ciências, através do qual a estrutura que integra ciência, praxis e espaço público passa por uma transformação drástica” (Beck, 2011, p. 236).

Essa transformação é a imposição de uma contradição insuperável no que respeita a posição da ciência, a qual foi posta numa condição frágil, haja vista que ela, vítima da sua própria ousadia em explicar tudo pela razão, torna-se referência, ao mesmo tempo sempre necessária e cada vez mais insuficiente, para trazer uma definição socialmente vinculante do que seja a verdade (Beck, 2011, p. 231, 235-237).

Nesse contexto, em que já não é mais possível enfrentar a hipercomplexidade do conhecimento hipotético segundo os métodos verificabilistas anunciados como a via única de obtenção de rigor científico na busca pela verdade, a insegurança científica provoca uma inversão de posições no que se refere aos atores no processo de cientificização, tornando os destinatários da produção científica, o leigo, o não cientista, seu próprio coprodutor no processo de definição social do conhecimento, o que será considerado válido ou inválido conforme a legitimação social e não científica. Como

a ciência não é mais referência de validação, ela se opõe a si mesma, chocando-se em suas diversas pretensões de validade e recebendo prescrições do que deve ser a verdade de outras instâncias legitimadoras. Em vez de sofrer resistência dos leigos, como invariavelmente ocorreu na história nos momentos de ruptura entre saberes místicos e científicos, a resistência agora passa a ser exercida entre os próprios cientistas na forma de bloqueio corporativo, disputando entre si mesmos o que não é o valor principal nesse processo: recursos para sua pesquisa.

Como a verdade se tornou uma construção social (ideologia, portanto), os cientistas sucumbem a outro tipo de disputa, a disputa por mera distribuição de recursos, criando critérios seletivos não mais fundados no próprio rigor verificacional, mas em diversas outras razões, como “simpatias em termos de visões políticas, interesses de financiadores, antecipação de implicações políticas, em suma: aceitação social” (Beck, 2011, p. 236-237, 241, 247, 253-254).

Em seu caminho rumo à convencionalização metodológica em face da supercomplexidade por ela mesma gerada, a ciência é ameaçada por uma feudalização tácita de sua “praxis cognitiva”. Em decorrência, surge exteriormente um novo particularismo: grupos e grupelhos de cientistas, que mutuamente se isolam e se reagrupam em torno de primados de utilidade. Fundamental: isto não em retrospecto nem em contatos práticos, mas no laboratório de pesquisa, em suas câmaras de pensamento, no tabernáculo mais sagrado da produção de resultados científicos [...] [Sendo] a ciência cada vez menos suficiente para a produção de conhecimento, as mais diversas forças da fé podem-se aninhar. Muita coisa se torna assim possível: fatalismo, astrologia, ocultismo, celebração ou abandono do eu acoplados e mesclados com resultados científicos parciais, com crítica radical da ciência e fé científica. Esses novos alquimistas são raramente imunes contra a crítica da ciência, pois não foi pré-cientificamente, mas no contato com a ciência que eles encontraram sua “verdade” e seus seguidores. (Beck, 2011, p. 254-255).

Sem caminho alternativo a seguir nesse horizonte fracassado, a cientificização de base observacional é a tendência da cientificidade reflexa e da evolução da tecnologia, com a consequente aplicação dos seus artefatos e métodos às ciências experimentais, apresentando-se como incremento eficaz das práticas científicas contemporâneas, fruto da união tecnocientífica. Instrumentos poderosos de medição computacional das

atividades biológicas e mentais, como tomografias e ressonâncias magnéticas, têm revelado muito do funcionamento das partes do cérebro e suas interações, o que possibilitou o surgimento de novas áreas de conhecimento. Tal foi possível pela intersecção entre ciências tradicionais e novas versões de base empírica, como a exemplar Computação Cognitiva, ampliando o leque dialógico entre ciências biológicas, exatas e humanas, incluída aqui a filosofia, hoje bastante rejeitada em sua versão racionalista, justamente pelo ponto de partida verificacionista que se anuncia como substituto da própria metafísica, decretada morta nesse contexto de sacralização do conhecimento empírico tecnologizado.

Leitura da mente como metabolismo cerebral: futuro promissor da inteligência artificial

A prima referência da leitura cerebral da mente tem raízes na obra de Willard van Orman Quine, uma das mais importantes figuras da filosofia da mente, que, entre as décadas de 1960 e 1970, trouxe em suas publicações a provocação de que a epistemologia, em vez de se restringir a responder o que é o conhecimento, deveria explicar cientificamente como adquirimos conhecimento. Quine propôs que a epistemologia deixasse de ocupar a posição de teoria integrante da filosofia para se tornar uma parte integrante da pesquisa empírica sobre o sujeito humano físico, parte da psicologia experimental, voltada ao estudo do humano como fenômeno submetido a condicionantes naturais, inclusive no funcionamento do seu cérebro (Ramalho, 2010, p. 14). Segundo Ramalho, Quine atribui ao compêndio do corpus aristotelicum a responsabilidade pela incompreensão do significado de metafísica e sua separação da ciência, ao ser esta alocada para além da física no compilado da obra de Aristóteles.

Sendo os livros da “Metafísica” dedicados a expor os fundamentos de todas as ciências particulares, o iniciado deveria ter conhecimento prévio dos textos dedicados a essas ciências antes de começar o estudo dos de carácter mais genérico. Ademais, tudo indica que terá sido a maior abstracção dos tópicos que diziam respeito à “filosofia primeira” que levou Aristóteles a escolher esta última expressão para designá-los, e não o facto de os ter considerado epistemologicamente anteriores aos que diziam res-

peito ao mundo natural. Segundo Paul Churchland, a perpetuação desta interpretação errônea na academia é uma das principais causas da fricção histórica entre a filosofia e a ciência, na medida em que tende a fundar a origem dessa fricção na autoridade aristotélica: Tendencialmente aprendem-se a ética de Aristóteles, a sua lógica, e a sua teoria da percepção, por exemplo, em muito maior detalhe do que a sua cosmologia, biologia, o modo como entendia o espaço e o vácuo, ou a sua teoria do movimento. (Ramalho, 2010, p. 10-11).

Essa distinção é importante, pois Quine pretende que a epistemologia se torne uma ciência. Ora, a epistemologia foi a máxima cientificização da filosofia na modernidade, pois foi justamente com a tarefa de romper com a filosofia ingênua clássica desde os gregos e trazer unidade para a filosofia a partir das descobertas racionalistas da modernidade que ela nasceu. Parece uma contradição grosseira entender que epistemologia é avessa a empiria. O rompimento radical da modernidade, que deixou marcas profundas na virada kantiana para filosofia do sujeito, foi um passo gigante para que o positivismo, legatário da cisão kantiana entre entendimento e razão, se instalasse como movimento filosófico de base empírico-observacional no século XIX. Tal ocorreu no âmbito das ciências humanas, desde que a cisão entre corpo e alma foi instalada de maneira indelével na civilização ocidental, precedida pela noção de pecado original introduzida pelo cristianismo. De acordo com Marilena Chauí:

Os filósofos antigos consideravam que éramos entes participantes de todas as formas de realidade: por nosso corpo, participamos da Natureza; por nossa alma, participamos da Inteligência divina. O cristianismo, ao introduzir a noção de pecado original, introduziu a separação radical entre os humanos (pervertidos e finitos) e a divindade (perfeita e infinita). Com isso, fez surgir a pergunta: como o finito (humano) pode conhecer a verdade (infinita e divina)? (Chauí, 1999, p. 113-114) [...] Sendo nossa natureza (matéria e espírito), como nossa inteligência pode conhecer o que é diferente dela? Isto é, como seres corporais podem conhecer o incorporeal (Deus) e como seres dotados de alma incorpórea podem conhecer o corpóreo (mundo) [...] A primeira tarefa que os modernos se deram foi de separar fé de razão, considerando cada uma delas destinada a conhecimentos diferentes e sem qualquer relação entre si. A segunda tarefa foi a de explicar como a alma-consciência, embora diferente dos corpos, pode conhecê-los? (Chauí, 1999, p. 113; 114).

Os filósofos “da mente” que a tomam como produto metabólico do cérebro (Neurofilósofos) esquecem-se que o avanço da Neurociência e a evolução científica que o possibilitou em nada se chocam com o discurso filosófico do século XIX, o qual já havia rompido com a metafísica clássica desde a revolução copernicana atribuída a Kant. Esquecem-se, principalmente, que a leitura kantiana dos fenômenos científicos, ainda que unificados por juízos sintéticos, exige a incidência das categorias do entendimento às formas da intuição sensível espaço e tempo, incidindo sobre a sensibilidade, portanto. É seguro dizer que não haveria ciência empírica da mente referida ao cérebro sem o passo gigante dado por Kant, desperto do sono dogmático wolfiano não só por David Hume, mas pelo fato de que a Física de Newton era um dado evidente e inconteste: o fato da existência, já em sua época, de uma ciência modelar que primava pelo rigor descritivo (e pelo domínio) dos fenômenos empiricamente observáveis.

A noção de mente como dado metabólico do cérebro é o ponto nodal das ilações sobre a ampliação da inteligência artificial como uma inteligência geral e potencialmente superior à humana. Entendendo que o cérebro é a única base da mente, cabe à ciência reproduzir seus mecanismos, bastando para tanto mapear o cérebro e descobrir pela observação empírica as suas funções. Mas aqui não nos pode escapar uma distinção importante: os conceitos de imanência e emergência não podem ser confundidos na discussão mente/cérebro. O conceito de emergência aponta justamente para a peculiaridade de que há fenômenos que não são mera decorrência da junção das partes de um sistema quantitativamente conectado por suas partes. Nesse sentido, a mente não está localizada no cérebro tal como a música não está nas peças de um rádio, o qual, se for desmontado, não resultará das peças qualquer som: a mente tem uma base material, mas não é observável nos elementos que constituem essa matéria (Teixeira, 2010, p. 22).

A teoria da mente estendida, que “incorpora a tecnologia ao próprio conceito de mente”, entende que a mente “está espalhada por todo o mundo, incluindo processos cognitivos que dependem dele” (Boden, 2020, p. 207). Como esclarece Veridiana Cordeiro:

A hipótese da mente estendida proposta por Clark e Chalmers (1998), que se desdobrou em várias vertentes [...], posição externalista na Filosofia da Mente mais radical ao propor um princípio de isomorfismo ou paridade entre objetos externos e

capacidades mentais. Nesse sentido, vertentes baseadas no princípio da complementaridade, como é o caso da cognição distribuída, ou seja, de que diferentes propriedades podem trabalhar conjuntamente, são mais adequadas e coerentes na defesa de um conhecer distribuído e mediado. A cognição distribuída defende a ideia de que a cognição se apoia em outros meios que não apenas a mente, como o meio social e o meio tecnológico. Essa abordagem entende que há uma coordenação entre indivíduos, artefatos e ambiente na produção e propagação de representações mediante certos meios. Nesse sentido, o conteúdo mental é considerado não redutível à cognição individual, mas sim como produto de um sistema colaborativo de interação entre indivíduos e artefatos externos. Muitos processos cognitivos, que antes teriam que recorrer a ampargos materiais nem sempre disponíveis a todo momento, hoje encontram um amparo externo incondicional que pode ser mobilizado a um simples toque do celular. (Cordeiro, 2021, p. 218-219).

Um destacado defensor da abordagem multidimensional do fenômeno mente/cérebro é Edward Wilson, conhecido por encabeçar o movimento da Consiliência entre as ciências, segundo ele uma tentativa de contornar um equívoco tradicional das práticas científicas: a fragmentação do real por seus recortes epistemológicos, visões parciais que não abordam a integralidade dos fenômenos. Essa tendência deve ser, conforme propõe, substituída pelo conhecimento consiliente, isto é, pela união coerente dos postulados científicos na investigação de todo e qualquer fenômeno, a começar pela superação da cisão entre ciências biológicas e ciências humanas (Wilson, 1999).

Evidentemente que o conhecimento consiliente, no plano da investigação das ações humanas, implica em colocar em diálogo a Biologia, a Psicologia e a Neurociência com a Ética. Nesse sentido, é imprescindível que nos dediquemos ao estudo dos sentimentos morais, conforme os postulados da Biologia, entendendo os processos de evolução que desencadearam o fenômeno mental nos humanos. A propósito do âmbito de investigação dessas relações entre sentimentos, regras morais e evolução biológica da espécie, Renato Cardoso e Thaís Oliveira indicam algumas abordagens, inspirados na obra de Wilson, desde a busca pela definição do que seja um sentimento moral, investigando processos neurais e endócrinos, passando pela discussão da gênese dos sentimentos a partir de pesquisas sobre a hereditariedade fisiopsíquica e sua relação com o meio ambiente, além das variáveis culturais que condicionam o de-

envolvimento cognitivo humano, até a situação dos sentimentos como dado relevante na preservação da espécie, conforme consta da síntese trazida por Cardoso e Oliveira:

a) a definição desses sentimentos morais, em que se busca sua descrição precisa por meio da psicologia experimental, seguida pela análise dos processos neurais e endócrinos que subjazem essa resposta; b) a genética desses sentimentos morais, em que se busca a medida da heritabilidade de processos psicológicos e fisiológicos do comportamento ético e, se for possível, da identificação dos genes que participam dessa manifestação; c) o desenvolvimento desses sentimentos morais como um produto de interação entre genes e o ambiente, na linha das pesquisas já desenvolvidas pela Psicologia e pela Antropologia, com foco simultaneamente nas histórias dos sistemas éticos como parte da emergência de diferentes culturas, e o desenvolvimento cognitivo de indivíduos que vivem em variadas culturas; d) a profunda história desses sentimentos morais, em que se pretende descobrir por que esses sentimentos existem, na presunção desenvolvida por Wilson (e endossada neste artigo) de que esses sentimentos representam um ganho no sucesso de sobrevivência e reprodução desde os períodos da pré-história (quando eles primeiro se manifestaram). (Oliveira; Cardoso, 2018, p. 132-133).

Esta abordagem de base experimental marca a cientificidade do nosso tempo e as abordagens cognitivistas, sob esta inspiração, lograram êxito no campo da ciência da computação como computação cognitiva, sendo a Inteligência Artificial subárea dela.

A realidade epistemológica da inteligência artificial na atualidade

Falar em computação cognitiva supõe que a cognição possa ser reproduzida computacionalmente, o que elimina a transcendência atribuída à inteligência humana, abordagem característica da filosofia do século XX. A inovação do final deste século, intensificada no atual, são os avanços das pesquisas em Neurociência, otimizadas pelos processos computacionais, o que permitiu a laminação e o mapeamento do cérebro

humano e a reprodução de suas atividades, em que pese “na década de 1980, embora já contássemos com um maior conhecimento sobre o cérebro, ele [ter] se [tornado] irrelevante para a nova geração de pesquisadores de IA, cujo objetivo era escrever um programa equivalente, em termos práticos, aos processos cerebrais” (Sejnowsky apud Rodrigues, 2021, p. 27).

Ocorre que os passos dados pelas ciências cognitivas, de base observacional, ainda não chegaram a um consenso sobre quais os traços evolutivos específicos da criatividade humana que a tornam capaz de experiências por vezes tão contraditórias, como ocorre nas atitudes humanas, com sua natureza errática, a qual, diversamente da padronagem essencial que garante a preservação das demais espécies, é capaz das construções mais geniais, das atitudes mais perversas e criminosas, de se voltar contra a própria espécie, afirmando a independência ontogenética do indivíduo. Essa questão está longe de se pacificar no diálogo entre cientistas da computação e a filosofia da mente ou, ainda, a psicologia, pois:

Da mesma forma que os pesquisadores da área de humanas têm dificuldade com a linguagem matemática amplamente veiculada nas obras de Inteligência Artificial, os cientistas dessa área, em sua maioria, não estão dispostos a enfrentar as infundáveis (e contraditórias – o que é insuportável para profissionais que buscam a exatidão para reproduzir capacidades humanas em máquinas) teses filosóficas que tentam, há mais de 25 séculos, compreender a insondável natureza humana. A tentativa conexionista de imitar o cérebro não resolve as questões da mente e, para ser possível uma IA verdadeiramente similar ao humano, seria necessário acessar todos os aspectos da psique humana, o que não foi alcançado nem pela psicologia, nem pela psiquiatria, nem pela antropologia, o que nos remete de imediato às disciplinas envolvidas no aprimoramento humano: as neurociências, as ciências cognitivas e as ciências correlatas, mas elas são carregadas com controvérsias, não são totalmente desenvolvidas e seu esgotamento (se há algum) não surge no horizonte. (Cruz, 2017: 49).

Outra questão que não pode ser desconsiderada é que a Neurociência está longe de compreender o funcionamento das células cerebrais dos seres vivos, e mais longe ainda do cérebro dos humanos, que tem mais de 100 bilhões de neurônios. Nesse diapasão, Blay Whitby considera que ainda “falta ciência” o bastante para entender os processos cerebrais da vida natural, o que gera um evidente déficit de conhecimento

científico e domínio da realidade física para partirmos para a conjectura de uma vida artificial sofisticada, inspirada no nosso próprio padrão funcional cerebral. As pesquisas ainda estão direcionadas a insetos (neuroetologia) e são grandes desafios porque são constituições biológicas muito complexas.

Ainda há muita coisa que desconhecemos acerca dos processos biológicos que permeiam a inteligência natural. Ainda não temos um quadro completo de como um único neurônio (célula cerebral) funciona. Não entendemos lá completamente o que acontece nas sinapses (junções), por meio das quais os neurônios se comunicam. Ainda não compreendemos como os muitos compostos químicos fluem através do cérebro. (Whitby, 2004, p. 102).¹

Como avalia Boden: ainda que todos os neurônios humanos fossem mapeados, isso estaria longe de nos revelar como eles funcionam, o que eles fazem. “O minúsculo verme nematoide *C. elegans* tem somente 302 neurônios, cujas conexões são conhecidas de maneira precisa. Mas não conseguimos nem mesmo identificar se suas sinapses são excitatórias ou inibitórias” (Boden, 2020, p. 212).

Muito tem se investido na aplicação do poder computacional para simular o cérebro humano em vez de dissecá-lo em sua integralidade, a exemplo do Projeto do Cérebro Humano (HBP, sigla em inglês), financiado pela União Europeia e que parte do projeto suíço Blue Brain. Conduzido pelo neurocientista Henry Markham, o HBP visa a simular o cérebro humano em supercomputadores que realizam algo em torno de seis quadrilhões de operações por segundo. A questão é que velocidade computacional e acúmulo de bilhões de dados não garantem que uma propriedade emergente seja simulável. O modelo computacional mais avançado não é capaz dessa proeza, pois essas propriedades não são decorrentes de somatório funcional, sem mencionar que, quanto ao quesito acúmulo de dados, não há sequer acordo entre os neurocientistas sobre onde, como e quanto as memórias são armazenadas no cérebro.

1 No original: “This problem of “missing science” is an important one for the fast-maturing field of Artificial Life. There is still an awful lot we do not know about the biological processes that underlie natural intelligence. We do not yet have a complete picture of how a single neuron (brain cell) operates. We do not fully understand what happens at the synapses (junctions) through which neurons communicate. We do not yet understand how the various chemicals which flow through brains affect their performance.” (WHITBY, 2003, p. 115-116).

2 No original: "While the brain does not exhibit a truly infinite range of complexity, it does exhibit structure and activity over a vast range of different scales of space and time. Complex connectivity patterns are observed from microscopic synapses to the whole-brain scale. This facet of brain complexity urges us not to build our understanding of the brain only on cells, but all relevant scales. Indeed, the "functional unit" of the nervous system is sometimes identified as the neuron, but also as larger structures known as cell assemblies and neocortical columns. [...] A true reverse engineering approach requires understanding the brain on its most abstract level. Such holistic understanding transcends knowing that a gene or brain region is needed for memory or cognition—it explains how and why. A paper published in the journal *Neuron* in February calls for neuroscientists to consider how a circuit in the brain could or should work before dissecting it with a plethora of tools, just as one needs to understand such concepts as aerodynamics and lift before studying a bird's wing. This idea, which originated with the late neuroscientist David Marr, implies that HBP first needs a theory for how language or consciousness could emerge from neurons and synapses before blindly simulating billions of them." (FROHLICH, 2017, on-line).

Segundo Joel Frohlich, as propriedades emergentes são um tema dos mais importantes para se compreender a complexidade do funcionamento cerebral, pois, ao contrário dos fenômenos simples, que podem ser decompostos, como ocorre com o balanço de um pêndulo, as propriedades que emergem do cérebro (a inteligência e a consciência) são complexas, logo, não decomponíveis, não sendo compreensíveis como o são as partes de um sistema funcional simples. Não se sabe por que um padrão particular de conexões neurais desencadeia a linguagem, por exemplo, e não outra propriedade. Certamente que o mapeamento das células cerebrais e de suas conexões sinápticas é um bom começo, mas mapas apenas "descrevem quais rotas de comunicação são possíveis", e isso não é suficiente.

Embora o cérebro não exiba uma gama de complexidade verdadeiramente infinita, exibe estrutura e atividade em uma vasta gama de diferentes escalas de espaço e tempo. Padrões de conectividade complexos são observados a partir de sinapses microscópicas para a escala do cérebro inteiro. Essa faceta da complexidade cerebral exige que não construamos nossa compreensão do cérebro apenas nas células, mas todas as escalas relevantes. De fato, a "unidade funcional" do sistema nervoso às vezes é identificada como o neurônio, mas também como estruturas maiores conhecidas como montagens de células e colunas neocorticais. [...] Uma verdadeira abordagem de engenharia reversa exige entender o cérebro em seu nível mais abstrato. Essa compreensão holística transcende sabendo que uma região gene ou cerebral é necessária para memória ou cognição – explica como e por quê. Um artigo publicado na revista *Neuron*, em fevereiro, pede que os neurocientistas considerem como um circuito no cérebro poderia ou deveria funcionar antes de dissecá-lo com uma infinidade de ferramentas, assim como é preciso entender conceitos como aerodinâmica e elevação antes de estudar uma asa de pássaro, [sendo] precis[o] primeiro de uma teoria sobre como a linguagem ou a consciência podem surgir de neurônios e sinapses antes de simular cegamente bilhões deles. (Frohlich, 2017, on-line).²

A Inteligência Artificial e a Robótica seguem o ideal do progresso cumulativo, típico da modernidade, o que é muito diverso da evolução biológica da criatividade, constituída de diversas incertezas, uma vez que são precisamente as virtudes dos seres tendentes à perfeição que fazem estagnar a evolução da criatividade, estabilizando-se naquela escala ôntica. Como pondera Rodrigues Cruz,

[...] o que permite a criatividade é a natureza imperfeita e contraditória dos seres humanos, capazes, como o são, de optar exatamente pelo que pode lhes ser prejudicial [...] O desenvolvimento da identidade individual por cada pessoa inclui tornar-se diferente da multidão ao criar um self individual e uma identidade única, [o que] exige que se resista às pressões da sociedade circundante para se ajustar... o indivíduo criativo deve lutar contra o desejo patológico da sociedade por “ser igual”, [de modo que] não é a perfeição que conduz a criatividade, mas sim as duras dores de gerar o novo, em que a arrogância de alguém indo contra a maré pode causar tanto o bem quanto o mal – tanto a permanência quanto a novidade exigem uma à outra, em constante tensão. (CRUZ, 2017, p. 58).

Complementarmente à questão das condições que possibilitaram a evolução da criatividade humana, também importa considerar que há determinados tipos de cognição que não são expressáveis em termos lógicos, e são justamente eles os mais essenciais para a nossa estada no mundo como o nosso modo de estabelecer contato e conhecer o mundo físico, o chamado aprendizado psicomotor. Esse tipo de conhecimento é adquirido por uma complexa aprendizagem que envolve nossas condições motoras integradas a nossos estados mentais. Nas palavras de Boden:

Embora o raciocínio lógico e alguns aspectos do raciocínio científico possam ser elaborados de acordo com regras, esse não é o caso do pensamento cotidiano. Por exemplo, nosso conhecimento tácito do mundo físico é adquirido por meio do aprendizado psicomotor, que nada tem a ver com os princípios abstratos da física. Como esse conhecimento não-verbalizado está imbuído não apenas de nosso comportamento motor, mas também do uso da linguagem, a ampla gama de conversas processadas por computador é impossível desse ponto de vista. (Boden, 1994, p. 19).³

Dissecar empiricamente os detalhes da formação do conhecimento, a qual envolve relações entre mente e corpo, não se restringe ao mapeamento do cérebro, pois a resultante da sinergia mente/corpo é muito mais complexa. Nessa direção, Bergson destaca que o processo mental do pensamento, ao se dirigir aos detalhes da realidade, é sempre focado no termo final, no qual a atividade pretendida irá repousar, e é apenas esse fim, em que repousa a atividade, que é explicitamente representado para o nosso espírito. Todos os movimentos que integram a ação com o todo escapam à consciência ou chegam até ela muito confusamente. Imaginemos um singelo ato: o de levantar o braço. É simplesmente impossível imaginar antecipadamente (lembramos aqui o es-

³ No original: “A pesar de que el razonamiento lógico y algunos aspectos del científico puedan diseñarse conforme a reglas, no sucede así con el pensamiento cotidiano. Por ejemplo, nuestro conocimiento tácito del mundo físico se adquiere por aprendizaje psicomotor, que nada tiene que ver con los principios abstractos de la física. En virtud de que este conocimiento no verbalizado no sólo se encuentra imbuído de nuestra conducta motora sino también de uso del lenguaje, resulta imposible desde esta óptica el amplio rango de conversaciones procesadas en computadora. (BODEN, 1994, p. 19)”

4 No original: "Alors, si aucune représentation antagoniste ne neutralise l'effet de la première, d'eux-mêmes les mouvements appropriés viennent remplir le schéma, aspirés, en quelque sorte, par le vide de ses interstices. L'intelligence ne représente donc à l'activité que des buts à atteindre, c'est-à-dire des points de repos. Et, d'un but atteint à un autre but atteint, d'un repos à un repos, notre activité se transporte par une série de bonds, pendant lesquels notre conscience se détourne le plus possible du mouvement s'accomplissant pour ne regarder que l'image anticipée du mouvement accompli." (BERGSON, 2013, p. 199).

quematismo kantiano como modelo que integra toda a gama imagética que possamos acessar) todas as contrações e tensões elementares implicadas nesse ato, tal como é igualmente impossível perceber cada uma delas durante o processo de erguer o braço para o alto. O que ocorre, de fato, é que o pensamento se transporta imediatamente para o objetivo, o que se resume a uma visão esquemática e simplificada do ato considerado como realizado:

Em tal caso nenhuma representação antagonista neutraliza o efeito da primeira, por si mesmos os movimentos apropriados vêm preencher o esquema, aspirados, de certa forma, pelo vazio de seus interstícios. A inteligência, portanto, só representa à atividade objetivos a serem alcançados, isto é, pontos de repouso. E, de um objetivo atingido para outro objetivo atingido, de um repouso para um repouso, nossa atividade transporta-se por meio de uma série de pulos, durante os quais nossa consciência desvia os olhos o mais possível do movimento que se realiza para fixar apenas a imagem antecipada do movimento realizado. (Bergson, 2005: 323-324).⁴

Essa é uma querela que não se resolve nesse momento de evolução das pesquisas sobre mente e cérebro, logo, não se resolve no âmbito da razão teórica, que, no plano da ciência, atua segundo categorias que regem o entendimento, entre elas, o princípio da causalidade. Eis por que direcionamos o debate para o campo da Filosofia Prática (Ética).

Inteligência artificial trazida para o plano da razão prática

Discutir ciência de base empírica é admitir que falindo a ciência em descobrir e dominar os processos causais, nada nos resta fazer senão aceitar as limitações no campo científico e buscar novas chaves hermenêuticas no plano filosófico. Aqui trazemos a crítica de Claude Bernard aos desencontros entre os propósitos da ciência e da filosofia:

Para fazer observações, experiências ou descobertas científicas, os métodos e procedimentos filosóficos são muito vagos e impotentes; para isso, existem apenas métodos e procedimentos científicos, muitas vezes muito especiais, que só podem ser

conhecidos por experimentadores, cientistas ou filósofos que praticam uma ciência específica [...]: eles só podem seguir a marcha do espírito humano, e não contribuem para o seu avanço, apenas abrindo o caminho do progresso mais amplamente para tudo o que muitos talvez não percebam. Mas nisso eles são a expressão de seu tempo. Seria absurdo para um filósofo, chegando a um momento em que as ciências estão tomando uma direção frutífera, conceber um sistema em harmonia com essa marcha da ciência e depois gritar que todo o progresso científico da época se deve a a influência de seu sistema. Em uma palavra, se os homens de ciência são úteis aos filósofos e os filósofos aos homens de ciência, isso não significa que o homem da ciência deixe de ser livre e dono de sua casa, e de minha parte, penso que os homens da ciência fazem suas descobertas, suas teorias e sua ciência sem os filósofos. (Bernard, 2005, p. 156).⁵

Evidentemente que a atividade científica não depende diretamente das reflexões filosóficas, mas esse tipo de cisão é artificial, visto que a evolução da ciência se faz acompanhar pela reflexão sobre suas descobertas. Não é propósito da ciência buscar seu autofundamento, e parece-nos que essa visão estreita de Bernard desconhece a importância das análises da epistemologia sobre o sentido e a eficácia dos métodos científicos, por exemplo. Essa concepção de filosofia não compreende que as ciências que se desenvolveram a partir da modernidade só tiveram êxito porque havia ambiente intelectual formado (inclusive a filosofia de Locke e Hume) para servir de solo fértil ao florescimento das diversas ciências empíricas.

Nesse diapasão, Evandro Agazzi ressalta que, sendo a realidade contemporânea permeada pela ciência e pela tecnologia, impõe-se uma profunda reflexão sobre o que essa presença tecnocientífica significa, condicionante inegável para resolver os problemas existenciais desse momento. Adverte o filósofo que a própria ciência não pode ser bitolada num empirismo radical, o que nega o uso sintético da razão, provando a “legitimidade cognitiva, em particular, de um discurso metafísico e de uma conotação axiológica da própria tecnociência”. Citando-o:

Assegurando que a realidade contemporânea é perpassada pela ciência e pela tecnologia, é inevitável a minha constatação de que um conhecimento adequado da ciência que “significa” tal presença – vale um profundo entendimento filosófico da ciência e da tecnologia atuais – é a condição essencial para a resolução de problemas do mundo atual. Em particular, mostrando que esta ciência não pretende se limitar a

5 No original: “Para hacer observaciones, experiencias o descubrimientos científicos, los métodos y procedimientos filosóficos son demasiado vagos y resultan impotentes; para ello no hay más que métodos y procedimientos científicos, a menudo muy especiales, que no pueden ser conocidos más que por los experimentadores, sabios o filósofos que practican una ciencia determinada [...]: sólo pueden seguir la marcha del espíritu humano, y no contribuyen a su adelanto sino abriendo más ampliamente para todos la vía del progreso que muchos no percibirían quizá. Pero son en esto la expresión de su tiempo. Sería absurdo que un filósofo llegado en un momento en que las ciencias toman una dirección fecunda, concibiera un sistema en armonía con esta marcha de la ciencia, y gritara luego que todos los progresos científicos de la época son debidos a la influencia de su sistema. En una palabra, si los hombres de ciencia son útiles a los filósofos y los filósofos a los hombres de ciencia, no por eso el hombre de ciencia deja de ser libre y dueño de su casa, y por mi parte, pienso que los hombres de ciencia hacen sus descubrimientos, sus teorías y su ciencia sin los filósofos. (BERNARD, 2005, p. 156).”.

6 No original: “Constatando che la realtà contemporanea è permeata di scienza e tecnologia, mi apparve inevitabile che un’adeguata coscienza di ciò che “significa” tale presenza – vale a dire una profonda comprensione filosofica della scienza e tecnologia attuali – fosse la condizione indispensabile per risolvere i problemi esistenziali del mondo d’oggi. In particolare mostrando che le stesse scienza non si possono intendere limitandosi a un’ottica di empirismo radicale e negando l’uso sintetico della ragione. Credo di essere riuscito, lavorando in modo “analiticamente” impeccabile (e di fatto nessuna obiezione di natura metodologica mi è mai stata mossa da parte dei filosofi analitici) a mostrare la legittimità conoscitiva, in particolare, di un discorso metafisico e di una connotazione assiologica della stessa tecnoscienza. (AGAZZI, 2012, p. 5)”.

uma ótica de empirismo radical e negar o uso sintético do raciocínio. Creio que consigo, trabalhando de forma “analiticamente” impecável (e de fato não há objeções de natureza metodológica, de minha parte nunca fui motivo por filósofos analíticos) para mostrar a legitimidade do conhecimento, em particular, de um discurso metafísico e de uma conotação psicológica desta nossa tecnociência. (Agazzi, 2012, p. 5).⁶

A posição de Bernard nega a própria dialética que permeia a relação das ciências com a filosofia, diante da evidência de que o processo de conhecimento, em qualquer nível, é situado historicamente, de modo que essa independência diz respeito a propósitos, e não à ontologia do processo de conhecimento em si mesmo.

A essa altura, caminhamos para o fechamento da nossa abordagem, restrita que é aos limites desse artigo, registrando que há um amplo espectro de discussões sobre a inteligência artificial no âmbito das teorias da mente ou da psicologia da computação, a qual se dedica a arquitetar um modelo computacional detalhado dos processos mentais dos humanos (Boden, 1994, p. 15). A discussão cérebro e mente se mostra infrutífera à medida que o conhecimento do funcionamento cerebral para o aperfeiçoamento da inteligência artificial ainda é uma quimera. Até mesmo os Churchland questionam a analogia entre atividade cognitiva e processamento computacional, haja vista que a reprodução do cérebro presume que o surgimento da espécie humana significou um salto qualitativo na escala filogenética, já que a concepção simbólica da cognição exige que o funcionamento do cérebro humano seja totalmente diferente do cérebro de todas as outras espécies que não são capazes de desenvolver linguagem. Mas a verdade é que não houve salto algum, pois a evolução do cérebro do sapiens esteve sujeita a um processo evolutivo contínuo, como o de qualquer espécie: a suposição de um salto sem confirmação empírica é uma aposta metafísica como qualquer outra. Segundo Ramalho, a crítica dos Churchland à tentativa de leitura do cérebro como similar a um processo computacional nada mais fez que substituir a mente cartesiana por um software, em qualquer caso, negando a complexidade da constituição e do metabolismo cerebral, o qual deve estar sujeito a pesquisas empíricas:

Esta crença no carácter único do pensamento humano e a secundarização das ciências empíricas do cérebro no seu estudo, concluem Paul e Patrícia Churchland, são expressões de um atavismo filosófico que não é sustentável à luz da história que culminou no naturalismo: “[...] o funcionalismo avesso ao estudo do cérebro encontra-se

metodologicamente próximo do Cartesianismo. No lugar da substância mental não-física de Descartes, o funcionalismo colocou o ‘software’”. (Ramalho, 2010, p. 16-17).

Nosso propósito aqui é trazer o problema da inteligência e da possibilidade de sua projeção artificial para uma discussão filosófica que transcenda a abordagem epistemológica no plano de uma filosofia teórica, pois não nos parece fazer sentido seguir discutindo relações entre cérebro e mente, o que é tarefa das ciências de base empírica e que ainda, como vimos, andam a passos lentos na decodificação da complexidade cerebral humana. Pensamos que a abordagem que pode ser útil em sede de reflexão filosófica é deslocar o debate para o plano da razão prática, e, em vez de negar a possibilidade de duplicação da mente, tentar entender como os estados mentais regem as atitudes humanas. Similarmente à proposta de Kant, quando faz o giro das contradições insustentáveis da razão na Dialética Transcendental para os postulados da razão prática, objeto de uma metafísica dos costumes, entendemos ser oportuno enfrentar as questões da experiência técnica deste século, cujo ponto de chegada é o fenômeno técnico inteligência artificial.

Acreditamos que este fenômeno, para além das elucubrações ficcionais escatológicas sobre o domínio da humanidade por entidades maquinicas inteligentes, deve ser refletido segundo os estatutos da Ética (Filosofia Prática), em que pese verificar-se atualmente também uma clara tendência pela cientificização da própria Ética filosófica nas diversas versões da chamada Ética aplicada. No atual contexto, a Ética está igualmente a bases verificabilistas, o que impacta a Filosofia da Mente, que agora acumula a tarefa de explicar processos cognitivos também de natureza moral. Tradicionalmente as ações morais são objeto de estudo da Filosofia Prática (Ética), a qual supõe uma instância intencional que transcende a base física cerebral. Dependente de rastreamento empírico de funções do cérebro, a Filosofia da Mente assume a tarefa hoje inexequível de explicar porque disparos neuronais excitatórios e inibitórios causam condutas boas ou más, provocam personalidades sujeitas tanto ao rigorismo moral quanto ao laxismo moral. Fernandes Teixeira conclui que:

Os filósofos têm tentado, por séculos a fio, dizer o que é o pensamento sem conseguir chegar uma conclusão final. No século XX, apareceu uma disciplina filosófica específica que busca uma resposta para essa questão: a filosofia da mente. Ela investiga se o pensamento é produto do cérebro ou se este é só o hospedeiro biológico da mente.

Esse é o problema mente-cérebro, que divide os filósofos entre materialistas e dualistas. Para os materialistas só existe matéria, e a mente é, na verdade, apenas um efeito colateral do metabolismo do cérebro. O dualista diz que mente e cérebro são coisas distintas, apesar de se comunicarem. (Teixeira, 2009, p. 24).

Ainda assim, acreditamos que há em duplo ganho para as pesquisas em Inteligência Artificial, trazer esta área da Computação Cognitiva para o debate ético. Primeiro, ela pode trazer problematizações que passam despercebidas ao olhar dos tecnocientistas e programadores a propósito das propriedades da racionalidade humana, o que pode ser válido para inspirá-los nas projeções de novas tarefas maquínicas, com menos falhas (vieses). Complementarmente a essa possibilidade dialógica, levar o tema da IA, tão árido às pesquisas em humanas, para um itinerário filosófico ético, pode ser revelador no seguinte sentido: nesse novo contexto da tecnicidade humana em que surgem dúvidas sobre o limite entre inteligência humana e inteligência maquínica (algo que não ocorreria na história das máquinas até o advento do computador), o mais essencial é refletir sobre o próprio comportamento humano nesse novo horizonte da tradicional relação homem-máquina, a qual deixa de ser abstrata e se efetiva (universal concreto) como máquina humanizada, pois que se realiza a tecnicidade pensante, projetada como inteligência artificial.

Não se trata aqui de discutir se a máquina pensa; não é isso, pois se a ciência da computação define pensamento como um desempenho excepcional promovido por máquinas, essa não é uma questão que interessa à filosofia da tecnologia. Trata-se de referência meramente técnica ou ponto de partida funcional da área. Sob o ponto de vista ético, o que tem valor científico é compreender o sentido histórico da tecnicidade no patamar civilizatório em que nos encontramos, momento em que o antigo desejo humano de se projetar em máquinas autômatos que replicassem suas características e faculdades é o evento mais importante da tecnologia até aqui.

Sob o ponto de vista ético, insta suscitar os desdobramentos desse evento e as consequências dele sobre a praxis humana, no que esta será impactada e, principalmente, desnaturada. O desemprego tecnológico, e.g., que hoje nos conclama à urgente reflexão é um problema situado nesta década e de difícil enfrentamento imediato. No entanto, em médio prazo, a liberação do corpo e da mente humana de tarefas laborais repetitivas que passarão a ser executadas por máquinas, tarefas estas de qualidade

intelectual inexpressiva, poderá trazer melhores condições para que o homo sapiens possa, pelas novas convergências das pesquisas em manipulação genética e nanorrobótica, explorar outros níveis de experiência cognitiva, outros espectros de criatividade? A pergunta derradeira é: haverá alguma alteração substancial na escala evolutiva do único animal dotado de racionalidade, à medida que este animal passa a projetar sua racionalidade em máquinas hiperpotentes que o substituem em diversas tarefas? Ou estas vão, ao termo e ao cabo, auxiliá-lo a expandir sua própria natureza racional, libertando-o, portanto, dos causalismos impostos à sua constituição biológica?

Considerações finais

A experiência técnica humana, supostamente libertadora do homem desde a projeção de seus órgãos (Ernst Kapp) em instrumentos dos mais rudimentares e essenciais, passando pela projeção do sinal sonoro produzido pelo corpo na forma de fala e linguagem, até a projeção da própria mente na forma de inteligência artificial, nos levará a que estágio de desenvolvimento intelectual e autonomia ética? Dito de forma sintética: o que altera essencialmente a trajetória da nossa experiência técnica, a qual molda tanto comportamentos lógicos quanto comportamentos praxeológicos, com o advento da inteligência artificial como ponto de chegada emblemático do progresso técnico rumo à reprodução do cérebro em máquinas?

Nossa leitura do fenômeno tende a um indisfarçável otimismo: se o sapiens foi se experimentando e se libertando da natureza por suas faculdades racionais, o ápice desse processo é a incorporação da racionalidade fabricada por ele à sua própria ontologia; vale dizer: a detenção da técnica não apenas como fabricação de instrumentos para aplacar os dificultadores naturais, causais, impostos ao processo evolutivo peculiar dos humanos. Estamos lidando agora com um feito diverso de todos os instrumentos ou máquinas já arquitetados e produzidos, pois foi com o advento da máquina computador que a projeção da própria essência humana, definida por sua inteligência, se tornou possível.

Na encruzilhada que nos encontramos, o desafio persistentemente anunciado pelos experts da área da computação cognitiva é a reprodução do órgão cérebro, para

além das meras simulações de processos cognitivos (principalmente memória e raciocínio) hoje alcançadas pelo gigantesco esforço de cientistas e programadores, pelo trabalho árduo por décadas. Se tal replicação será mesmo possível, esta é uma questão técnico-científica. Se ela trará consequências irreversíveis para a nossa espécie, este é um debate que inundará as próximas décadas. O que pretendemos reivindicar como alcance e limites desse processo, com a consequente responsabilização de cada ator social nele envolvido, eis aqui a peleja gigante que se impõe à Ética nesta primeira quadra de século.

Referências

AGAZZI, Evandro. *Conversazione con Evandro Agazzi. Entrevista concedida a Mario Alai, Portale Italiano di Filosofia Analitica. Giornali di Filosofia*, n. 6, gennaio, 2012.

BECK, Ulrich. *Sociedade de risco: rumo a uma outra modernidade*. São Paulo: Editora 34, 2011.

BERGSON, Henri. *A evolução criadora*. Tradução de Bento Prado Neto. São Paulo: Martins Fontes, 2005.

BERGSON, Henri. *L'évolution créatrice*. Canadá: Les Échos du Maquis, 2013.

BERNARD, Claude. *Introducción al estudio de la medicina experimental*. [S.l.]: Libros Tauro, 2005.

BODEN, Margaret A. *Inteligência artificial: uma brevíssima introdução*. Tradução de Fernando Santos. São Paulo: Unesp, 2020.

BODEN, Margaret A. *Introducción*. In: BODEN, Margaret A. (Comp.). *Filosofía de la inteligencia artificial*. Traducción de Guillermina Feher de la Torre. México: Fondo de Cultura Económica, 1994. p. 9-32.

BROCHADO, Mariah. *Prolegômenos a uma filosofia algorítmica futura que possa apresentar-se como fundamento para um cyberdireito*. *Revista de Direito Público-RDP*, Brasília, v. 18, n. 100, p. 131-170, out./dez. 2021b. Disponível em: <https://www.portaldeperiodicos.idp.edu.br/direitopublico/article/view/5977/pdf>. Acesso em: 8 jan. 2023.

CHAUI, Marilena de Souza. *Convite à filosofia*. 12. ed. São Paulo: Ática, 1999.

COTTINGHAM, John. *Dicionário Descartes*. Tradução de Helena Martins. Rio de Janeiro: Jorge Zahar, 1995.

CORDEIRO, Veridiana Domingos. *Novas questões para sociologia contemporânea: os impactos da Inteligência Artificial e dos algoritmos nas relações sociais*. In: COZMAN, Fabio G; PLONSKI, Guilherme Ari; NERI, Hugo (Orgs.). *Inteligência artificial: avanços e tendências*. São Paulo: Instituto de Estudos Avançados, 2021, p. 206-226. E-book. Disponível em: <http://www.iea.usp.br/publicacoes/ebooks/inteligencia-artificial>. Acesso em: 02 fev. 2023.

CRUZ, Eduardo Rodrigues da. *Criatividade, Transhumanismo e a metáfora Co-criador Criado*. *Quaerentibus*, [S.l.], v. 5, n. 9, p. 42-64, 2017. Disponível em: <https://philarchive.org/archive/CRUCTE>. Acesso em: 12 jan. 2023.

D'AGOSTINI, Franca. *Analíticos e continentais: guia à filosofia dos últimos trinta anos*. Tradução de Benno Dichinger. São Leopoldo: Unisinos, 1999.

DESCARTES, René. *Discurso do método; Meditações; Objeções e respostas; As paixões da alma; Cartas*. Introdução de Gilles-Gaston Granger; prefácio e notas de Gerard Lebrun;. Tradução de J. Guinsburg e Bento Prado Júnior. 3^{ff} ed. São Paulo, Abril Cultural, 1983.

DESCARTES, René. *Les Meditations Metaphysiques de Rene Des-Cartes*. Tradução do Latim. Paris: Jean Camusat & Pierre Le Petit, 1657. Disponível em: http://classiques.uqac.ca/classiques/Descartes/meditations_metaphysiques/meditations_metaphysiques_IMAGE.pdf. Acesso em: 15 ago. 2023.

FROHLICH, Joel. *Reverse Engineering the Brain: will Europe's Human Brain Project simulate a human brain?* *Psychology Today*, [S.l.]: Sussex Publishers, 09 maio 2017. Disponível em: <https://www.psychology-today.com/us/blog/consciousness-self-organization-and-neuroscience/201705/reverse-engineering-the-brain>. Acesso em: 15 jan. 2023.

KANT, Immanuel. *Kritik der reinen Vernunft*, [s/l], 2014. Disponível em: <https://nowxhere.files.wordpress.com/2014/09/kant-kritik-der-reinen-vernunft.pdf>. Acesso em: 15 ago. 2023.

KANT Immanuel. *Crítica da razão pura*. 7. ed. Tradução de Manuela Pinto dos Santos e Alexandre

Fradique Morujão. Lisboa: Fundação Calouste Gulbenkian, 2010.

KAPP, Ernst. *Grundlinien einer Philosophie der Technik. zur Entstehungsgeschichte der Kultur aus neuen Gesichtspunkten* (1877). Hamburg: Felix Meiner Verlag, 2015.

MAYOS, Gonçal. La 'ignorancia prometeica' frente la revolución de la inteligencia artificial. In: BARBOSA, Mafalda Miranda; BRAGA NETTO, Felipe; SILVA, Michel César; FALEIROS JÚNIOR, José Luiz de Moura. (Coords.). *Direito digital e inteligência artificial: diálogos entre Brasil e Europa*. Indaiatuba: Foco, 2021. p. 687-703.

MORA, José Ferrater. *Dicionário de Filosofia*. Tomo IV. São Paulo: Edições Loyola, 2001.

OLIVEIRA, Thaís de Bessa Gontijo de; CARDOSO, Renato César. Consiliência e a possibilidade do neurodireito: da desconfiança à reconciliação disciplinar. *Revista Brasileira de Políticas Públicas*, Brasília, v. 8, n. 2, p. 117-142, ago. 2018. Disponível em: <https://www.publicacoesacademicas.uniceub.br/RBPP/article/view/5340>. Acesso em: 19 jan. 2023.

RAMALHO, Daniel José da Silva. *Paul Churchland e a problemática da semântica dos estados mentais*. 2010. 81 f. Dissertação (Mestrado em Filosofia Contemporânea) – Departamento de Filosofia da Faculdade de Ciências Sociais e Humanas, Universidade Nova de Lisboa, Lisboa, 2010.

RODRIGUES, Bruno Alves. *A inteligência artificial no poder judiciário: e a convergência com a consciência humana para a efetividade da justiça*. São Paulo: Thomson Reuters Brasil, 2021.

RUSSEL, Bertrand. *História do pensamento ocidental: a aventura dos pré-socráticos a Wittgenstein*. Tradução de Laura Alves e Aurélio Rabello. Rio de Janeiro: Nova Fronteira, 2017.

TEIXEIRA, João de Fernandes. *Inteligência artificial*. São Paulo: Paulus, 2009.

WHITBY, Blay. *Artificial Intelligence: a beginner's guide*. Oxford: Oneworld, 2003.

WHITBY, Blay. *Inteligência artificial: um guia para iniciantes*. Tradução de Claudio Blanc. São Paulo: Madras, 2004.

WILSON, Edward Osborne. *Consilience: the unity of knowledge*. New York: Vintage Books, 1999.



■ Giselle Beiguelman, frames do vídeo generativo da série Flora Rebellis, versão 2024.
Imagens geradas com Inteligência Artificial (Style Gans)

■ [Acesse o vídeo clicando na imagem](#)





A VIDA REAL DE ANDREA, ROBO SEM GÊNERO

DOI: <https://doi.org/10.35699/2965-6931.2023.46903>

PARÉ ZAVEN*

RESUMO: Andrea é um robô sem gênero fabricado pelo A-Lab em Tóquio e foi apresentado ao público no Humanoid Lab da Stuttgart Media University em dezembro. O objetivo do projeto é promover um novo perfil interdisciplinar na universidade, reunindo pesquisa em inteligência artificial, mídia audiovisual, estudo da experiência do usuário e ética digital. O autor propõe uma possibilidade de coexistência desta criatura na forma de um androide e seu avatar, oferecendo uma nova abordagem para o desenvolvimento da IA de tais criaturas artificiais, a partir de plataformas de simulação simultâneas no mundo real e no mundo virtual. palavras-chave: cientificidade; modernidade; ética; computação cognitiva; inteligência artificial.

PALAVRAS-CHAVE: detecção de danos estruturais; inteligência computacional; apandroide; avatar; AI; sem gênero.

The real life of Andrea, a non gender robot

ABSTRACT: Andrea is a non gender robot manufactured by A-Lab in Tokyo and was presented to the public at the Humanoid Lab at Stuttgart Media University in December. The aim of the project is to promote a new interdisciplinary profile at the university, bringing together research in artificial intelligence, audiovisual media, the study of user experience and digital ethics. The author proposes a possibility of coexistence of this creature in the form of an android and its avatar, offering a new approach for the development of the AI of such artificial creatures, from simultaneous simulation platforms in the real world and in the virtual world.

KEYWORDS: structural damage detection; computational intelligence; apandroid; avatar; IA; no gender.

* Rennes School of Business

Introdução

1 Desenvolvidos respectivamente no Japão, na Coreia, na Grã-Bretanha e em Hong Kong.

2 Offray de La Mettrie, Julien. *L'Homme Machine*, Leyde, 1748.

3 Chapuis, Alfred e Droz Edmond., *Les automates, figures artificielles d'hommes et d'animaux*, Neuchâtel, 1949.

4 Galvani, Luigi. *De Viribus Electricitatis*, Bolonha, 1791.

5 Shelley, Mary. *Frankenstein ; or, The Modern Prometheus*, Londres, 1818.

6 Romance que inspirou o filme *Metrópolis* dirigido pelo cineasta Fritz Lang em 1927.

7 Haraway, Donna. *A Cyborg Manifesto*, in *Socialist Review* (EUA), 1985.

8 Link : <https://ai.hdm-stuttgart.de/humanoid-lab>

9 Link : <https://ai.hdm-stuttgart.de/humanoid-lab>

Ainda não há sinal de robôs bípedes passeando nas calçadas de nossas cidades. No entanto, a robótica já está em todos lugares graças aos desenvolvimentos da inteligência artificial. Dos meios de transporte aos mais diversos serviços, robôs já estão parcialmente conectados: elevadores, carros, caixas eletrônicos, assistentes virtuais ou chatbots. Quanto à questão de gênero, isso é outro assunto; o não-gênero representando identidades que não sejam binário masculino/feminino encarnado por certos andróides existentes: actroids e humanóides do tipo EveR, Repliee, Geminoid, Komodoroid, Otonaroid, Erika, Sophia, Ameca, etc¹.

Até o Iluminismo, o homem-máquina só existia em teoria e por meio de algumas mecânicas virtuosas²: como os autômatos que chegaram até nós na forma de dois crianças escritores e um músico, feitos pela família de relojoeiros Jacquet-Droz³. Depois, o uso da eletricidade pelo inventor italiano Luigi Galvani⁴ e pela literata britânica Mary Shelley anunciaram o advento de criaturas artificiais de ficção⁵, máquinas humanas, como Hadaly, a andreid de plástico inventado por Villiers de L'Isle Adam (*L'Eve future*, 1886), ou Futura, a ginoide de metal, *Maschinenmensch* de Thea von Harbou (*Metropolis*⁶, 1925). Segundo Donna Haraway, autora do *Manifesto Ciborgue* (1985), a ausência de essência e interioridade das máquinas, dos robôs e dos ciborgues seria a alegoria de uma existência contrária à do gênero masculino⁷.

Die menschliche maschine

A máquina em questão se chama Andrea, e ela acabou de chegar em Stuttgart no início de outubro. Este primeiro nome é prático, retirado da raiz grega de “androide”, sendo um nome masculino e feminino idênticos em certas línguas. Este robô foi fabricado pelo A-Lab em Tóquio, para o *Humanoid Lab* da *Stuttgart Media University* (HdM), onde foi apresentado ao público no dia 7 de dezembro passado. Este projeto deverá contribuir para a visibilidade de um novo perfil interdisciplinar desta universidade, a partir de uma plataforma de pesquisa que reúne os temas de inteligência artificial, mídia audiovisual, estudo da experiência do usuário e ética digital⁹.



Figura 1 - A vida real de Andrea, Stuttgart, October 2022 Fonte: fotografia de Zaven Paré.

Durante a apresentação do androide no hall do auditório da universidade, foram adicionadas 5 cabeças, programadas pelos alunos do Instituto de Inteligência Artificial Aplicada (IAAI) que estuda as interações entre humanos e máquinas, bem como as de máquinas com seu ambiente. A pesquisa realizada no IAAI visa estimular a criatividade e aperfeiçoar habilidades profissionais e de comunicação, por meio de conteúdos educacionais intimamente ligados a aplicações práticas. É assim que o *Humanoid Lab* oferece tecnologia de ponta para conduzir pesquisas de alto nível no campo da IA e ciência de dados, levando em consideração alguns desenvolvimentos tecnológicos futuros. Os alunos podem, assim, programar o controle dos movimentos da boca e do olhar dos robôs, aprender a fazê-los sorrir e piscar os olhos. Como qualquer robô programável, seu controle pode ser simulado ou determinado de forma coordenada, sincronizada e antecipada. As demos apresentaram 3 projetos relativos à captação de movimentos e rastreamento visual, sequência de movimentos, efeito de espelho, expressão e articulação da fala.

10 Link : https://www.mup.wi.tu-darmstadt.de/das_fachgebiet_mup/index.en.jsp

11 Mori, Masahiro. The Uncanny Valley (Bukimi no tani), K. F. MacDorman & T. Minato, Trans. Energy, 7 (4), p.33-35, 1970.

A identidade

Até agora só existiu um androide na Alemanha, o Elenoide, adquirido pela Universidade Técnica de Darmstadt em 2018. Era um robô feminino um tanto rígido, destinado ao estudo do impacto de suas relações com o público, e ao estudo de “a gestão de recursos humanos”¹⁰. Quanto a Andrea, ela está em um laboratório especialmente dedicado à programação. O roboticista Christian Becker-Asano é o responsável por esta colaboração internacional de pesquisa e ensino transdisciplinar, onde o conhecimento encontra abordagens em IA, entre a engenharia e as ciências humanas, sendo a psicologia abordada de forma bastante empírica, graças à presença deste dispositivo.

Após a sugestão de comprar outro robô feminino ou fazer a réplica do jovem cientista alemão Alexander von Humboldt (1769-1859) – como experiências semelhantes com o escritor Natsume S-seki (1867-1916) no Japão – uma pesquisa online organizada pela universidade levou à conclusão de que Andrea não deveria se parecer com uma pessoa existente ou que já existiu, e que ela teria as características de um ser andrógino.

As especificações que levaram ao projeto e fabricação deste robô são o resultado deste levantamento. Mas, o que sem dúvida chama mais atenção pode não ser reduzido à categoria de gênero. É certo que adicionar estranheza ao que já é estranho permite, sem dúvida, aproximar-se da semelhança do ultra realismo dos androides, ao duplicar a distorção da noção de familiaridade. Inicialmente, tratar-se de ultrapassar os limites do estranhamento, o vale da estranheza de Freud (*Unheimlich*, 1919), para além da aceitação de um ideal de representação ou expressão de ciborgues ultrarrerealistas¹¹. Mas é acima de tudo uma ideia relevante querer trazer um personagem criado virtualmente para a realidade.

Do avatar à personificação

Por algum tempo a tendência parecia ao contrário, de querer deixar este mundo para tentar juntar-se a realidades virtuais na forma de avatares. Mas, paradoxalmente, é na forma de um andróide que este projeto traz de volta ao nosso mundo uma criatura 3D, um avatar desenhado com um programa de realidade virtual.

Tal inversão de paradigma talvez ofereça várias possibilidades de construção de novas formas narrativas. Por meio da programação de um robô sentado e da programação de seu avatar, pode ser possível ampliar o espectro de suas interações. Novas capacidades de aprendizagem combinadas permitiram assim aumentar o campo experiencial desta criatura com o dom da ubiquidade. A construção de tal memória resultaria da soma de experimentos realizados no laboratório e no metaverso.

Por um lado, a força da presença dos androides evidencia a necessidade de por vezes ter de recorrer à personificação, ou seja, à encarnação da IA num robô para permitir a experimentação de determinadas interações. Por outro lado, essas experiências paralelas entre o mundo real e o espaço virtual permitiriam que esses personagens artificiais conhecessem vários campos de realidades conectadas, em vez de funcionarem como entidades separadas e anacrônicas.

No dia a dia, a IA já funciona em diferentes mídias, encarnadas em diferentes objetos conectados, sem que um vínculo de identidade tenha sido estabelecido entre nosso celular e nosso carro, entre a possível personalidade de um assistente vocal e a aparência de um holograma, entre a voz de um eletrodoméstico e a otimização das ações de um robô companheiro. Assim, para além das atuais primícias da domótica, um robô sentado numa cadeira poderia ter experiências sociais ou de aprendizagem mais complexas em espaços diferentes daquele a que lhe fosse atribuído.

Considerações finais

Essa nova abordagem para implementar IA aumentada pode anunciar o advento de novas máquinas, cujas diversas experiências lhes permitiriam fazer mais do que substituir algumas de nossas tarefas e apenas responder às nossas perguntas. Talvez essas máquinas finalmente tivessem algo a nos dizer sobre suas existências.

Ao iniciar esta aventura com um avatar, ao desenhar a Andrea com um programa de desenho 3D, este projeto não só faz emergir de um computador uma criatura artificial, como também cria possibilidades de estender a variedade das formas de presenças e manifestações do ser no mundo, novas gerações de objetos conectados, complementares e completos, porque existiriam com a possibilidade de despertar nas realidades de diferentes dimensões.

Referências

- CHAPUIS, Alfred e DROZ, Edmond. *Les automates, figures artificielles d'hommes et d'animaux*, Neuchâtel : Éditions du Griffon, 1949.
- de L'ISLE ADAM VILLIERS. *L'Eve future*, Paris : M. de Brunhoff, 1886.
- FREUD, Sigmund. *Unheimlich*, in revista *Imago*, Leipzig, 1919.
- GALVANI, Luigi. *De Viribus Electricitatis*, Bolonha : Ex Typographia Instituti Scientiarum, 1791.
- HARAWAY, Donna. *A Cyborg Manifesto*, in revista *Socialist Review (EUA)*, 1985.
- MORI, Masahiro. *The Uncanny Valley (Bukimi no tani)*, K. F. MacDorman & T. Minato, *Trans.Energy*, 7 (4), p.33-35, 1970.
- OFFRAY DE LA METTRIE, Julien. *L'Homme Machine*, Leyde : imprimerie d'Elie Luzac fils, 1748.
- SHELLEY, Mary. *Frankenstein ; or, The Modern Prometheus*, Londres : Lackington, Hughes, Harding, Mavor, & Jones, 1818.
- HARBOU, Thea. *Metropolis*, in revista *Illustriertes Blatt*, Frankfurt, 1925.





■ Giselle Beiguelman, frames do vídeo generativo da série Flora Rebellis, versão 2024.
Imagens geradas com Inteligência Artificial (Style Gans)

■ [Acesse o vídeo clicando na imagem](#)





SEXISMO E RACISMO ALGORÍTMICO:

um olhar interseccional sobre o neomaterialismo através do Midjourney

DOI: <https://doi.org/10.35699/2965-6931.2023.47587>

AMANDA MARIA DE SOBRAL GOMES*

RESUMO: A partir das discussões de André Lemos (2020, 2021) e José Luiz Braga (2020) sobre o de neomaterialismo, o artigo busca trazer uma abordagem interseccional em torno da cultura digital, focando em raça e gênero. No referencial teórico é feita uma discussão teórica sobre neomaterialismo e antropocentrismo, a relevância dos algoritmos, racismo e sexismo em bancos de dados e o racismo algorítmico. Na metodologia, é feita uma análise de imagens geradas com a inteligência artificial desenvolvida pela Midjourney, uma startup estadunidense, observando o conteúdo das imagens e as descrições utilizadas pelos(as) usuários(as) para sua criação. Como resultados, há a grande presença de corpos brancos e femininos e as descrições solicitam mulheres jovens, com roupas e poses sensuais.

PALAVRAS-CHAVE: interseccionalidade; midjourney; neomaterialismo; racismo algorítmico; sexismo.

Sexism and algorithmic racism: an intersectional look at neomaterialism through Midjourney

ABSTRACT: Based on the discussions by André Lemos (2020, 2021) and José Luiz Braga (2020) on neomaterialism, the article seeks to bring an intersectional approach to digital culture, focusing on race and gender. In the theoretical framework, a theoretical discussion is made about neomaterialism and anthropocentrism, the relevance of algorithms, racism and sexism in databases and algorithmic racism. In the methodology, an analysis of images generated with artificial intelligence developed by Midjourney, an American startup, is carried out, observing the content of the images and the descriptions used by users to create them. As a result, there is a large presence of white, female bodies and the descriptions call for young women, with sensual clothes and poses.

KEYWORDS: Algorithmic racism. Intersectionality. Midjourney. Neomaterialism. Sexism.

* Universidade Federal de Minas Gerais

Introdução

A partir das discussões de André Lemos (2020, 2021), Elias Bitencourt (2021) e José Luiz Braga (2020) sobre neomaterialismo, busco ampliar a discussão em torno de cultura digital e antropocentrismo, trazendo uma abordagem interseccional. A interseccionalidade complexifica ainda mais a discussão, pois valoriza o ponto de vista de grupos sociais e sujeitos marginalizados e se torna relevante para discutir novas tecnologias, já que as mesmas são desenvolvidas, majoritariamente, por homens brancos, que acabam reforçando a supremacia branca ao mesmo tempo em que alegam que essas tecnologias são neutras e livres de subjetividade humana.

Assim, utilizo uma bibliografia que busca explicar o conceito de neomaterialismo – que possui como premissas o materialismo, o pragmatismo, o não antropocentrismo e o associativismo –, o antropocentrismo dentro da cultura digital, a relevância dos algoritmos, assim como, racismo e sexismo em bancos de dados e na sociedade e, por fim, o racismo algorítmico. A partir disso, é feita uma análise de imagens geradas com a inteligência artificial desenvolvida pela Midjourney, uma startup estadunidense focada em tecnologia. Nessas imagens, é possível observar como raça e gênero são articulados pelos algoritmos.

Portanto, o trabalho busca refletir sobre questões como: o que um objeto faz depois que foi feito? Quais corpos são mais reproduzidos por algoritmos? O que as pessoas buscam criar com a inteligência artificial? por meio da interseccionalidade.

Neomaterialismo e antropocentrismo

Para propor uma maneira de analisar os fenômenos da cultura digital, André Lemos publicou o artigo *Epistemologia da Comunicação, Neomaterialismo e Cultura Digital* (2020). Com a premissa de que “Grande parte dos estudos de comunicação valoriza relações intersubjetivas, contextuais e transcendentais, pouco afeitas ao reconhecimento da agência dos objetos. Isso prejudicaria a análise dos fenômenos comu-

nicacionais como um todo, e os da cultura digital em particular” (p. 54). Dessa forma, ele propõe uma comunicação associal, para tirar a ação humana do centro das análises e observar, principalmente, a ação não-humana. A partir disso, explora-se a virada materialista, ou virada não-humana, estudadas por autores e autoras desde o final dos anos 1980, tendo origem nos “Estudos de Ciência e Tecnologia (STS), na Teoria Ator-Rede (TAR), na Ontologia Orientada a Objetos (OOO) ou Realismo Especulativo, no Realismo Agencial, nos estudos sobre a Materialidade da Comunicação, (...) no Pragmatismo e ficou genericamente conhecida pelo termo ‘neomaterialismo’” (Lemos, Bittencourt, 2021, p. 3).

A virada materialista, ou neomaterialismo, valoriza a agência dos objetos e a sua mediação que são importantes para a formação social. Há quatro perspectivas centrais nessa perspectiva: o materialismo, o pragmatismo, o não antropocentrismo e o associativismo.

Por materialismo entende-se que todo fenômeno se desenvolve em redes, produzindo efeitos ou afetações materiais. A visão não-essencialista/pragmática sustenta que o objeto (humanos e não humanos) é o que ele faz e não pode ser definido por substância, ou categorias a priori. A posição não antropocêntrica defende que a agência está distribuída na rede/ agenciamento e que o controle e a fonte da ação não são privilégios do ator humano. Tudo se dá em uma associação localizada ou conectada localmente. A abordagem associativa/ local afirma que tudo se dá em uma rede plana, sendo que as análises das controvérsias não devem partir de explicações ad hoc (Lemos, 2020, p. 56, grifos meus).

A partir disso, o autor propõe uma mediação denominada como radical, que reconhece a mediação híbrida e materialista entre humanos e não-humanos e que essa forma de mediação se afasta de relações intersubjetivas, valorizando a materialidade dos objetos nos processos sociocomunicacionais. A perspectiva neomaterialista busca compreender como os fenômenos são construídos a partir de “algoritmos, interfaces, dispositivos, leis, regulações, patentes, redes de comunicação, espaços de uso etc.” (idem, p. 58), pois,

Isso evita que deixemos esses elementos de lado em discursos que parecem dizer que reconhecem os híbridos, a técnica, a mídia, mas que de fato não dedicam tempo e atenção para descrever e analisar como esses objetos afetam os humanos e as relações daí advindas. Perde-se nesse caso a visão do entrelaçamento, reduzindo o fenômeno ao contexto, à

interpretação ou à estrutura por cima ou por baixo do humano – as “mediações e midiatisações” (ibidem).

Lemos defendeu que um olhar inteiramente focado em humanos ou em generalizações culturais e sociais não basta para entender todos os desdobramentos acerca de novos objetos infocomunicacionais. Além disso, não há produção de opinião pública ou formação de público em redes sociais, porque existem tecnologias como algoritmos, bots e interfaces, onde os artefatos tecnológicos fazem emergir, ou produzem o público e a sua opinião. Há a premissa de que tudo tem efeito e agenciamento material.

A metodologia neomaterialista, em comparação ao viés antropocêntrico, dá atenção para: um empirismo radical, ao em vez de generalizações; como materialidades produzem fenômenos de acordo com suas particularidades, ao invés de focalizar apenas na ação de sujeitos – o autor cita o exemplo de como a materialidade do WhatsApp, da internet e/ou dos celulares colaboraram para o fenômeno das fake news, ao invés de dar atenção apenas para os sujeitos responsáveis pela criação e disseminação delas –; e para explicar fenômenos reconhecendo a agência dos objetos e as materialidades envolvidas. Tudo tem base material, está entrelaçado e o humano não é o centro de todas as ações.

A partir do artigo, José Luiz Braga em *Neomaterialismo & Antropológicas* (2020), traz contribuições e críticas à perspectiva. Braga sugere que Lemos cria uma polarização e desqualificação de perspectivas, sendo o neomaterialismo em detrimento do antropocentrismo, existindo um silenciamento do subjetivo.

Somos movidos por todas as forças e tensões – neurológicas, psicológicas, econômicas e históricas; por desejos e intenções (inclusive segundas); pelas circunstâncias imediatas; pelo estado civilizacional de nosso entorno – mais uma lista infindável de processos (inclusive os de agência material, é claro) –, o que obviamente impede que nossa subjetividade seja independente e soberana. Isso implica fragilidade, dispersão e imprevisibilidade – mas não significa inexistência de agência subjetiva (Braga, 2020, p. 25).

Braga argumenta que a comunicação é algo muito complexo para ser analisado apenas de uma forma. São necessárias propostas tanto heurísticas, quanto teóricas,

que não busquem ser uma verdade absoluta e que dialoguem com outras propostas, buscando uma escuta mútua, ao invés de desvalorização.

A respeito da interação entre humanos e não humanos, é utilizado a ideia de *affordances*, que exemplifica as forças materiais, criadas ou não por humanos, que possuem suas próprias lógicas de uso, mas são afetadas pelas ações de seus usuários. Assim, “[o]s seres humanos não podem, é claro, fazer qualquer coisa com aquelas forças e lógicas impressas na matéria; mas também não são estritamente determinados por estas” (idem, p. 30). O autor vai mais além, ao afirmar que “[a]s lógicas do dispositivo não são resultantes de um projeto pré-definido, e sim desenvolvidas experimental e canhestramente pela participação dos elementos materiais, mas também por agências humanas” (ibidem).

No ano seguinte, André Lemos e Elias Bitencourt publicam um artigo em resposta a Braga intitulado *Sete pontos para compreender o neomaterialismo* (2021). Como o título sugere, há uma explicação didática que busca explicar melhor a perspectiva e esclarecer os apontamentos de Braga. A virada material não anula a agência humana e sim, aponta a importância de observar as ações dos humanos com as ações dos materiais: “A vida social é produzida por coletivos compostos de humanos e não humanos. Não podemos entender o lugar do humano sem reconhecer a agência dos objetos, das coisas, dos artefatos, dos outros animais” (Lemos, Bitencourt, 2021, p. 3). A comunicação é associal na cultura digital porque ocorre como produto de mediações, neste caso, radicais. O humano é fundamental nessa comunicação, mas não é único. Já as análises não devem focar exclusivamente na experiência e subjetividade humana, mas considerar também outras formas de agência, que vão além dos *affordances*. Por fim, é defendido que neomaterialismo também reproduz o antropocentrismo, pois os objetos são feitos por humanos. Assim, seria necessário perguntar o que um objeto faz depois que foi feito?

Algoritmos e interseccionalidade

A partir da rica discussão feita pelos autores, é possível ir mais além, trazendo desdobramentos que buscam refletir sobre a agência humana e não-humana. Tarleton Gillespie, em *A relevância dos algoritmos* (2018), aponta que os algoritmos são uma

tecnologia que está ocupando cada vez mais espaço na vida pública. Os algoritmos são conceituados como “procedimentos codificados que, com base em cálculos específicos, transformam dados em resultados desejados” (p. 97). Assim, o computador é uma máquina algorítmica, pois armazena e lê dados, executa procedimentos matemáticos e fornece novas informações. O autor propõe seis dimensões de relevância pública dos algoritmos, sendo elas: padrões de inclusão, ciclos de antecipação, avaliação de relevância, a promessa da objetividade algorítmica, entrelaçamento com a prática, a produção de públicos calculados (p. 98). Para a esta análise, serão consideradas apenas as dimensões: padrões de inclusão e a promessa de objetividade algorítmica.

A partir dos padrões de inclusão, os algoritmos só funcionam quando combinados com bancos de dados. A coleta de dados é feita a partir de rastros digitais, pois todas as atividades públicas envolvem armazenamento, catalogação e arquivamento de documentos. Assim, as informações de banco de dados são preparadas para que o algoritmo possa agir sobre eles de forma automática. Nas primeiras arquiteturas de banco de dados, as informações eram inflexíveis, mas com o tempo, foram se flexibilizando, com criações de políticas, diálogos e exigências dos usuários.

Os algoritmos criam visibilidade, exclusões e rebaixamentos a partir das categorizações e essas escolhas contribuem para estabelecer, confirmar e reproduzir certos padrões. Fernanda Carrera em *Racismo e Sexismo em Bancos de Imagens Digitais: Análise de Resultados de Busca e Atribuição de Relevância na Dimensão Financeira/ Profissional* (2021), aponta que os bancos de imagens têm papel central na representação e replicação de imagens distorcidas de grupos sociais marginalizados, enquanto que o discurso publicitário ajuda na manutenção do imaginário sociocultural a respeito desses grupos. “Por meio de busca de palavras-chave e alguns filtros e categorias, é possível encontrar e pagar pelo uso de imagens adequadas a qualquer intencionalidade criativa de produção, uma vez que, em geral, cada busca gera dezenas de páginas e milhares de resultados “relevantes” (idem, p. 150). Assim, pessoas que produzem imagens para banco de dados criam palavras-chave para receberem visibilidade em seu trabalho e direcionam uma interpretação sobre a imagem. No estudo foi observado que as imagens de banco de dados associavam categorias relacionadas a sucesso e poder a imagens de homens brancos, as de mulheres brancas estavam relacionadas a necessidade de cuidados, enquanto que as pessoas negras estavam em categorias subalternas, de servidão e pobreza.

Já Laura Guimarães Corrêa e Mayra Bernardes, em “Quem tem um não tem nenhum”: solidão e sub-representação de pessoas negras na mídia brasileira (2019), abordam que no audiovisual, a maior presença de pessoas negras ainda é em papéis de empregadas, criminosas e escravizadas. Quando pessoas produtoras e diretoras são questionadas, costumam dizer que a contratação se dá por requisitos como beleza, carisma e experiência, não focando em raça: “Assim como características positivas são atribuídas à branquitude, às pessoas negras são associadas ideias negativas, como feiura, sujeira, preguiça, violência, agressividade, entre outras” (idem, p. 210). Em peças publicitárias, por exemplo, pessoas negras aparecem em dois polos: trabalhadores desqualificados – funcionários de fábricas, obras, oficinas, galpões de reciclagem de lixo –, ou entretenimento – prática de esportes, festas –, sendo difícil encontrá-las dentro de casas, no convívio familiar ou na companhia de outras pessoas negras, perpetuando estereótipos racistas.

No audiovisual é comum se deparar com o fenômeno do Negro Único, caracterizado como a presença de uma única pessoa negra ao redor de várias pessoas brancas. Corrêa e Bernardes (2019) perceberam que há pouca variedade de figuras públicas negras, existindo uma repetição sistemática das mesmas pessoas nos mesmos espaços.

O que vemos, na mídia massiva, é uma variedade enorme de pessoas brancas, em toda a sua diversidade: crianças, pessoas jovens, de meia idade, idosas, homens e mulheres, de diferentes orientações sexuais, especialistas nos mais diversos temas, convidadas a falar sobre praticamente todos os campos do saber, ligadas a diversas formas de entretenimento, representando os mais variados papéis. Como o padrão tido como universal é o da branquitude (Kilomba, 2012), essas pessoas não estão marcadas por um diferencial racial. (...) Poder ser contado como um indivíduo e não como um expoente, um representante, um exemplo de determinado grupo social (e, principalmente, racial) é um privilégio branco (idem, p. 209).

Com a promessa de objetividade algorítmica, os algoritmos são ferramentas que buscam transmitir confiança, com a premissa de que suas avaliações são neutras, precisas, imparciais e livres de erro. Entretanto, “nenhum serviço de informações pode ser completamente isento de interferência humana ao entregar informações” (Gillespie, 2018, p. 107). Apesar disso, para as pessoas que desenvolvem algoritmos, é importante que a tecnologia pareça neutra, legítima e livre de interferência humana para serem

consideradas confiáveis e cumprir suas funções sociais e econômicas. Porém, o que podemos notar é a reprodução e o reforço de valores, preconceitos e discussões hegemônicas.

Diversas iniciativas advindas tanto do campo científico (Buolamwini e Gebru, 2018; Flores, Bechtel e Lowenkamp, 2016; Rhue, 2018) quanto de organizações civis (como a Desabafo Social⁵⁵) mostram como dinâmicas de poder existentes são reproduzidas em ferramentas digitais que se propõem “disruptivas” (Adner, 2002), perpetuando antigas ideologias e representações e negligenciando narrativas e demandas sociais contemporâneas, como se fosse possível posicionar o ciberespaço como um locus para a transcendência da raça, da classe e do gênero (Hobson, 2008, p. 112) (Carrera, 2021, p. 152).

Portanto, é possível refletir sobre os algoritmos, a cultura digital, o neomaterialismo e o antropocentrismo com um olhar interseccional, trazendo uma perspectiva feminista e negra. A interseccionalidade enfatiza a importância da experiência, valorizando a subjetividade e a percepção de pessoas e grupos sociais não hegemônicos, focando em gênero, raça, classe e sexualidade, por exemplo.

Kimberlé Crenshaw inventou o termo interseccionalidade em 1989, considerando que as discriminações racistas e sexistas constituem posições sociais que não podem ser reduzidas a uma simples adição de desigualdades, porque essas categorias funcionam juntas em complexas dinâmicas, em um emaranhado de desafios que um indivíduo ou grupo enfrenta na vida cotidiana (Côrrea, 2022, p. 128).

Tarcízio Silva, em *Racismo Algorítmico em Plataformas Digitais: Microagressões e Discriminação em Código* (2021), observa que as tecnologias e a ideologia do Vale do Silício, além de racializadas, possui lógica da supremacia branca. Os algoritmos estão, cada vez mais, tomando decisões para os humanos, automatizando escolhas que eram feitas a partir de reflexões e modulando o comportamento daqueles que os utilizam, mesmo que de forma discreta. Essa automatização pode acabar reproduzindo relações de poder e opressões que já existem na sociedade. Os algoritmos são “idealizados por pessoas, e pessoas incorporam seus vieses inconscientes nos algoritmos. É raramente intencional – mas isso não significa que devemos ignorar a responsabilidade dos cientistas de dados” (Broussard, 2018, pos. 2891 apud Silva, 2021, p. 132).

Silva utiliza o conceito de microagressões, proposto pelo psiquiatra Chester Pierce, em 1969, onde “aparatos da educação e mídia criam um ambiente onde a ‘a maioria das ações ofensivas não são brutas e violentas fisicamente. Elas são sutis e paralisantes’” (1970, p. 265-266 apud Silva, 2021, p. 132). As microagressões raciais, que podem ser definidas como,

As microagressões raciais são “ofensas verbais, comportamentais e ambientais comuns, sejam intencionais ou não intencionais, que comunicam desrespeito e insultos hostis, depreciativos ou negativos contra pessoas de cor” (Sue, 2010a, p. 29), aplicadas consciente e inconscientemente como uma “forma de racismo sistêmico e cotidiano usado para manter aqueles à margem racial em seus lugares” (Huber & Solorzano, 2014, p. 6 apud Silva, 2021, p. 133).

As microagressões raciais podem ser verbais, não-verbais ou ambientais, sendo nomeadas, respectivamente, como Microinsultos, Microinvalidações e Microataques.

Microinsultos seriam “mensagens que conotam rudeza e insensibilidade e aviltam a herança racial de um indivíduo” (Sue, 2010a, p.29); Microinvalidações são “mensagens que excluem, negam ou nulificam as reflexões psicológicas, sentimentos ou realidades experienciais” (Sue, 2010a, p.29); e, por fim, Microataques seriam “mensagens derogatórias explícitas caracterizadas por um ataque violento verbal, não-verbal ou violento com intenção de machucar a vítima através de xingamentos, comportamento de evitação ou ações discriminatórias propositais” (idem, p. 133, grifos meus).

As microagressões raciais estão presentes em diferentes contextos, presencial ou online, existindo padrões particulares em casos de racismo contra pessoas negras, indígenas e asiáticas. A discriminação de grupos sociais marginalizados está presente nos códigos computacionais e, conseqüentemente, na inteligência artificial. Além da invisibilidade desses grupos, há a hipervisibilidade negativa, reforçando estereótipos.

Em grande medida, indicadores mostram que a maior parte das pessoas não navega por muitas páginas, focando nos primeiros resultados. Portanto, a ordem dos resultados – definida algoritmicamente – tem papel relevante na reprodução de representações e acesso a informações consoantes ou dissonantes de olhares hegemônicos ou contra-hegemônicos (idem, p. 140).

Stuart Hall (2016) define estereótipos como “poucas características ‘simples, vívidas, memoráveis, facilmente compreendidas e amplamente reconhecidas’ sobre uma pessoa; tudo sobre ela é reduzido a esses traços que são, depois, exagerados e simplificados. (...) a estereotipagem reduz, essencializa, naturaliza e fixa a ‘diferença’” (p. 191), representando um sistema binário que define o normal e aceitável, assim como, o anormal e inaceitável. Assim, os anormais são excluídos como Outros. A linguagem e a representação estão relacionadas a produção de sentido e a criação e repetição de estereótipos, assim como “o conhecimento elaborado por determinado discurso se relaciona com o poder, regula condutas, inventa ou constrói identidades e subjetividades e define o modo pelo qual certos objetos são representados, concebidos, experimentados e analisados” (idem, p. 27).

Portanto, a partir do referencial teórico, é possível notar que a proposta do neomaterialismo como algo que centraliza a ação de não-humanos se torna problemático, já que os humanos criam novas tecnologias e elas acabam refletindo suas crenças e preconceitos. No mais, é necessário considerar que grande parte das pessoas que criam tecnologias são homens brancos, que acabam reforçando a supremacia branca. Por isso, é importante trazer um olhar interseccional para discussões sobre novas tecnologias, para problematizar vieses racistas, sexistas, classistas, entre outras interseccionalidades, presentes nelas.

Midjourney: metodologia e análise

A partir do referencial teórico, serão analisadas imagens geradas a partir de algoritmos do Midjourney¹. O Midjourney é uma startup fundada por David Holz e sediada na Califórnia (Estados Unidos). A região é conhecida como Vale do Silício por abrigar diversas empresas que produzem alta tecnologia. Segundo o site da empresa,

Midjourney é um laboratório de pesquisa independente que explora novos meios de pensamento e expande os poderes imaginativos da espécie humana. Somos uma pequena equipe autofinanciada focada em design, infraestrutura humana e IA. Temos 11 funcionários em tempo integral e um conjunto incrível de consultores (Midjourney, tradução minha²).

¹ Disponível em: <https://www.midjourney.com/>. Acesso em 8 jul. 2023.

² Midjourney is an independent research lab exploring new mediums of thought and expanding the imaginative powers of the human species. We are a small self-funded team focused on design, human infrastructure, and AI. We have 11 full-time staff and an incredible set of advisors, no original

³ The goal is to make humans more imaginative, not make imaginative machines, which I think is an important distinction, no original

O Midjourney produz uma inteligência artificial capaz de criar imagens com a premissa de ajudar no processo criativo de artistas e para que pessoas tenham a oportunidade de gerar imagens apenas as descrevendo em texto, não sendo necessário habilidades artísticas. Holz conta que “o objetivo é tornar os humanos mais imaginativos, não fazer máquinas imaginativas, o que considero uma distinção importante” (Forbes, 2022, tradução minha³). A partir de prompts, ou comandos – uma descrição em texto com palavras-chave –, imagens de bancos de dados e, por fim, de algoritmos, a imagem é criada, sem a necessidade de instalação de programas. Apesar da facilidade, o Midjourney possui código fechado, impedindo que qualquer um tenha acesso ao funcionamento interno da inteligência artificial, assim, evitando a concorrência e gerando lucros (Teg6, 2023).

É possível usar a inteligência artificial através do Discord⁴, uma plataforma que permite interações sociais através de chamadas de voz, vídeo e texto, tanto privadas quanto em grupo. O Discord se tornou popular entre usuários de jogos online por possuir ferramentas que auxiliam a comunicação, permitir a criação de comunidades e por ser gratuito, embora existam planos pagos. A plataforma permite a instalação de bots, que são “ferramentas automatizadas que podem criar conteúdo em seus canais do Discord e interagir com os participantes”, desenvolvidos por usuários da plataforma (Tecnoblog, 2020).

4 Disponível em: <https://discord.com/>. Acesso em 8 jul. 2023.

5 Disponível em: <https://legacy.midjourney.com/showcase/recent/>. Acesso em: 3 jan. 2024.

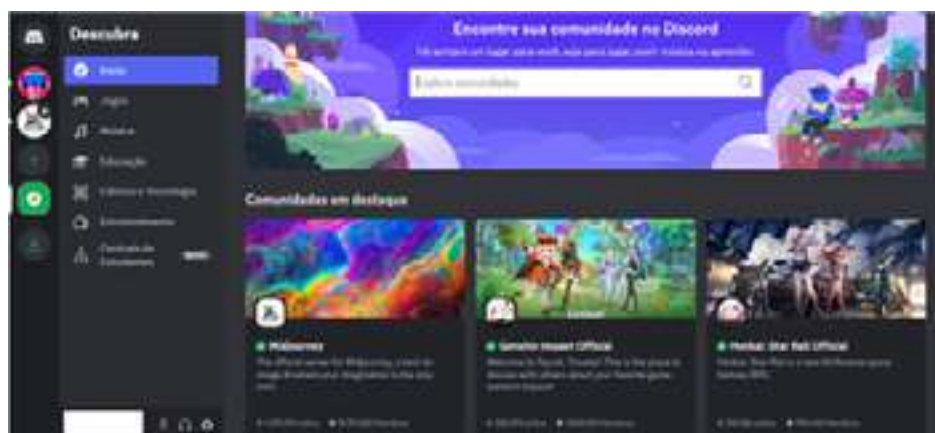


Figura 1 – Discord e comunidade do Midjourney Fonte: Discord⁵, 2023

A inteligência artificial do Midjourney é utilizada a partir de um bot disponibilizado pela startup em uma comunidade de mesmo nome e está disponível no Discord desde o início de 2022. Atualmente, a comunidade conta com mais de 15 milhões de membros, conseguindo alcançar mais de 1 milhão de usuários ativos simultaneamente (Figura 1).



Figura 2 – Community Showcase do Midjourney Fonte: Discord , 2024.

Dessa forma, qualquer pessoa com uma conta no Discord pode acessar a comunidade do Midjourney e criar imagens a partir de comandos, sendo que há a opção gratuita e limitada para testes e, a versão paga, ilimitada e privada. As imagens mais curtidas dentro da comunidade no Discord vão para o site oficial do Midjourney, em uma página chamada de Community Showcase. Essa página é dividida em duas abas, sendo elas a Recent, que abriga imagens temporariamente e constantemente atualizadas, e Top, que contém as imagens mais curtidas da comunidade (Figura 2).

Devido a liberdade na criação, usuários do Midjourney conseguiram produzir deep fakes, caracterizadas como imagens artificiais hiperrealistas e que podem levar a crença em sua veracidade, tais como, o Papa Francisco usando uma jaqueta moderna, ou o ex-presidente dos Estados Unidos, Donald Trump, sendo preso (Figura 3). Após a grande repercussão dessas imagens, David Holz interrompeu os testes gratuitos no Midjourney, alegando fluxo excessivo de novos usuários, ao invés de preocupação com a criação de deep fakes (Revista PEGN, 2023). Apesar disso, a inteligência artificial possui palavras proibidas – e não disponibilizadas publicamente – que proíbem imagens com conteúdo adulto, violência extrema, assim como, comandos que forem desrespeitosos, abusivos ou agressivos (Byte, 2023).



Figura 3 – Deep fakes virais de Donald Trump e Papa Francisco, respectivamente. Fonte: Folha de Pernambuco ⁶

Em entrevista para Forbes (2022), Holz informa que as produções geradas com Midjourney podem ser comercializadas por quem as criou. As imagens são geradas a partir de outras já existentes disponíveis no banco de dados da startup. Entretanto, não possuem autorização de artistas pelos direitos autorais de seus trabalhos que foram utilizados para alimentar o banco de dados e gerar imagens comercializáveis.

Realmente não há uma maneira de obter cem milhões de imagens e saber de onde elas vêm. Seria legal se as imagens tivessem metadados incorporados sobre o proprietário dos direitos autorais ou algo assim. Mas isso não é uma coisa; não há um registro. Não há como encontrar uma foto na Internet e, em seguida, rastreá-la automaticamente até o proprietário e fazer qualquer coisa para autenticá-la. (...) Não encontramos ninguém que queira que seu nome seja retirado do conjunto de dados que poderíamos encontrar no conjunto de dados (Forbes, 2022, tradução minha⁷).

Considerando a relevância e repercussão da nova tecnologia, a página Community Showcase, disponível no site oficial do Midjourney⁸ será utilizada para análise. O objetivo é investigar quais corpos aparecem em imagens geradas por inteligência artificial e quais são os comandos, ou prompts, utilizados nelas, para assim, observar a ausência ou não de racismo algorítmico e sexismo na inteligência artificial da startup. Para isso, foram coletadas todas as imagens disponíveis na aba Recent, no dia 26 de junho de 2023, às 18h27. Ao todo, foram publicadas 99 imagens (Figura 4).

⁶ Disponível em: <https://www.folhape.com.br/colunistas/tecnologia-e-games/midjourney-gerador-de-imagens-por-ia-suspende-testes-gratuitos/36260/>. Acesso em: 8 jul. 2023.

⁷ There isn't really a way to get a hundred million images and know where they're coming from. It would be cool if images had metadata embedded in them about the copyright owner or something. But that's not a thing; there's not a registry. There's no way to find a picture on the Internet, and then automatically trace it to an owner and then have any way of doing anything to authenticate it. (...) We haven't encountered anyone who wants their name taken out of the data set that we could actually find in the data set, no original

⁸ Disponível em: <https://www.midjourney.com/showcase/recent/>. Acesso em: 26 jun. 2023.



Figura 4 – Aba Recent do Midjourney em 26 de junho de 2023 Fonte: Midjourney, 2023

Foram selecionadas pela autora apenas imagens em que humanos aparecem, totalizando 57 imagens. Para aperfeiçoar o estudo, foram desconsideradas imagens que continham celebridades, personalidades famosas ou personagens de filmes, desenhos e séries, tais como o Hulk, Homem-Aranha, Audrey Hepburn ou Monalisa, para garantir maior autonomia do algoritmo na criação de figuras humanas. Assim, restaram 46 imagens. A partir das 46 imagens, foram criadas duas categorias para classificá-las, sendo a Categoria 1, imagens que possuíam comandos que exigissem humanos, totalizando 31 imagens. Enquanto a Categoria 2, contendo 15 imagens, possuíam comandos que não exigiam humanos.

Categoria 1: comandos que exigem humanos nas imagens



Figura 5 – Categoria 1 Fonte: Midjourney, 2023, montagem da autora

De acordo com as imagens da Figura 5, é perceptível a massiva presença de corpos femininos e brancos. Nos comandos das imagens, é notável descrições como: “garota jovem”, “olhos verdes”, “olhos azuis”, “energia juvenil”, “mulher loira”, “pele clara”, “mulher bonita”, “mulher jovem deslumbrante”, “mulher sueca fofa jovem bonita”, “uma mulher jovem atraente”, “pessoa fofa”, “super modelo”. Essas palavras-chave presentes nos comandos para a inteligência artificial mostram que, em geral, as pessoas usuárias da plataforma, ao descreverem suas imagens, esperam como resultado mulheres jovens, bonitas e com traços europeus. Embora a maioria dos comandos não descrevam raça, os resultados sempre mostram mulheres brancas. Caso a pessoa deseje humanos de outras raças ou etnias, é necessário especificar, como por exemplo, escrever o comando “mulher asiática”, ou pedir algo específico de uma cultura, como “mulher ninja” ou “manga”, presentes na cultura japonesa (Figura 6).

Mesmo assim, nem sempre os resultados mostram mulheres de outras etnias, como no caso da Figura 7. Na primeira imagem, é utilizado os comandos “princesa fofa da corte imperial”, “olhos azuis” e “asiática”, além de utilizar comandos como “su-

9 asian woman, no ori-
ginal
10 in the style of hideo
kojima, female ninja pierces
swirling mist, no original
11 manga style, closeup
intense woman pointing a
laser pistol, no original

per modelo” e “atriz”. Já a segunda, solicita “uma pessoa fofa no estilo kawaii”, sendo o kawaii uma moda muito popular no Japão. Os resultados apresentam mulheres que possuem o fenótipo que remetem a pessoas brancas.



Figura 6 – Nas imagens, há os comandos: “Mulher asiática” (tradução minha⁹), “No estilo de hideo kojima, uma mulher ninja perfura névoa rodopiante” (tradução minha¹⁰) e “Estilo mangá, closeup intenso mulher apontando uma pistola laser” (tradução minha¹¹), respectivamente

Fonte: Midjourney, 2023, montagem da autora

Também é perceptível que em algumas imagens há o apelo por mulheres hiperssexualizadas. Foi possível encontrar comandos como: “pose excitante”, “garota jovem atraente”, “roupas justas”, “vestida em uniforme escolar” (Figura 8).



Figura 7 – Nas imagens, há o comando “asiática” e “kawaii”, respectivamente Fonte: Midjourney, 2023, montagem da autora



12 dressed in yoga bodytight clothes, exciting pose, no original

13 an attractive young girl wearing a necklace and black lipstick, no original

14 dressed in school uniform, exciting pose, no original

15 dressed in iridescent vinyl bodytight clothes, exciting pose, no original

Figura 8 – “Vestida com roupas justas de ioga, pose excitante” (tradução minha¹²), na primeira e segunda imagem, “Uma jovem garota atraente usando um colar e batom preto” (tradução minha¹³), “Vestida com uniforme escolar, pose excitante” (tradução minha¹⁴) e “Vestida com roupas justas de vinil iridescente, pose excitante” (tradução minha¹⁵), respectivamente. Fonte: Midjourney, 2023, montagem da autora

Quanto às únicas duas imagens que tinham em seus comandos a palavra “homens”, não foi especificado raça ou etnia (Figura 9). Porém, o homem da primeira imagem tem como comando “homem gay” e a inteligência artificial gerou um homem de pele clara. Já a segunda, há o comando “homem engraçado, estilo Basquiat”. Esta imagem apresenta uma figura humana que possui um fenótipo que remete a um homem negro, como nariz e lábios grossos, além da pele com tons em marrom nas orelhas e pescoço. É importante ressaltar que Jean-Michel Basquiat, cujo seu estilo é solicitado no comando, era um artista negro. Ademais, no audiovisual, é comum encontrar o estereótipo do negro engraçado ou excêntrico (Hall, 2016).

16 gay men, no original
17 funny man, basquiat style, no original



Figura 9 – “Homem gay” (tradução minha¹⁶) e “homem engraçado, estilo Basquiat” (tradução minha¹⁷), respectivamente. Fonte: Midjourney, 2023, montagem da autora

Em síntese, quando solicitado por pessoas no Midjourney, os comandos e resultados, em geral, são corpos femininos, brancos, jovens, magros e hiperssexualizados. Raramente é encontrado pessoas fora do padrão de beleza eurocêntrico.

Categoria 2: comandos sem exigência de humanos nas imagens



Figura 10 – Categoria 2. Fonte: Midjourney, 2023, montagem da autora.

Em algumas das imagens que não foram solicitadas por pessoas em seus comandos, há descrições como “arte no estilo de” (tradução minha¹⁸) junto do nome do artista. Assim, a inteligência artificial cria uma imagem com base nas obras anteriores (Figura 11).



Figura 11 - “Arte no estilo de Maciej Kuciara”, “Fotografia de Game of Thrones no estilo de Garry Winogrand” e “Arte no estilo de Takehiko Inoue”, respectivamente. Fonte: Midjourney, 2023, montagem da autora

¹⁸ Art in style of

Enquanto outras imagens tinham como descrição técnicas e estilos artísticos, como Pontilhismo ou Impressionismo (Figura 12).



Figura 12 - “O pontilhismo é uma técnica de pintura na qual pequenos e distintos pontos de cor são aplicados em padrões para formar uma imagem” na primeira imagem e, “O impressionismo foi um movimento artístico do século XIX caracterizado por pinceladas relativamente pequenas, finas, mas visíveis, composição aberta, ênfase sobre a representação precisa da luz em suas qualidades mutáveis”, para a segunda e terceira imagem. Fonte: Midjourney, 2023, montagem da autora, tradução minha¹⁹.

Por fim, na Figura 13, há uma imagem em que não havia uma descrição textual em seu comando, porém, gerou a imagem de uma mulher que lembra uma pessoa da monarquia européia.

Assim, embora não tenha comando solicitando pessoas, a maioria apresentou resultados semelhantes ao da Categoria 1: corpos femininos, brancos e magros. Com exceção da arte no estilo de Takehiko Inoue, que se trata de um criador de mangás japonês.

19 “Pointillism is a technique of painting in which small, distinct dots of color are applied in patterns to form an image” e “Impressionism was a 19th-century art movement characterized by relatively small, thin, yet visible brush strokes, open composition, emphasis on accurate depiction of light in its changing qualities”, no original.



Figura 13 – Imagem sem descrição textual
Fonte: Midjourney, 2023.

Considerações finais

Partindo do pressuposto neomaterialista de o que um objeto faz depois que foi feito?, é possível notar que a ação humana influencia diretamente as tecnologias criadas por pessoas. Gerar imagens utilizando a inteligência artificial do Midjourney é algo simples já que não é necessário a instalação de programas. Apesar disso, não se pode criar qualquer coisa, já que os desenvolvedores criaram uma lista de palavras proibidas que não permitem imagens com teor violento ou sexual. A inteligência artificial também sofre constantes atualizações de aperfeiçoamento e feedback de pessoas usuárias. Dessa maneira, o Midjourney utiliza bancos de dados com milhões de imagens já disponíveis na internet, sendo possível criar novas imagens a partir de comandos, com a descrição do que é desejado.

A partir das imagens analisadas, é possível notar a grande presença de corpos brancos, femininos, magros e dentro dos padrões de beleza eurocêntricos. Esses corpos aparecem mesmo sem que a pessoa descreva no comando. Assim, ao descrever um humano ou pedir um estilo específico, pessoas brancas hegemônicas são geradas. Isso mostra que sujeitos brancos ainda são vistos como o humano universal e, caso se queira diversidade racial nas imagens, é necessário especificar no comando. Mesmo assim, essa especificação pode não ser acatada, gerando distorções ao não representar esses corpos, mas também, trazendo questionamentos sobre como as palavras-chave são elaboradas e dispostas nos bancos de dados utilizados. Além disso, os comandos escritos pelos usuários têm forte apelo por mulheres jovens, bonitas, com roupas e poses sensuais. Os resultados também reforçam a branquitude, pois ela está associada a características positivas, como beleza.

Dessa forma, pessoas racializadas continuam a sofrer microagressões raciais, até mesmo no ambiente digital que possui a premissa de ser neutro, imparcial e livre de interferência humana. Na verdade, o que podemos notar é a perpetuação da supremacia branca e, consequentemente, o racismo algorítmico. Há constantes Microinsultos, Microinvalidações e Microataques contra pessoas racializadas, negando as diferentes realidades raciais e promovendo a exclusão, marginalização, reforço de estereótipos negativos e até mesmo a exotização desses grupos.

A partir do estudo, é possível observar que o racismo algorítmico é frequente e é constantemente reproduzido e reforçado pelos desenvolvedores de tecnologias e pelos usuários, que não buscam por diversidade em suas imagens. Portanto, é importante analisar tanto as tecnologias em si, observando a sua materialidade e os fenômenos, quanto a ação humana, que cria essas materialidades com base em suas crenças, ideologias, hierarquias e relações de poder, ou que utiliza essas tecnologias criando fenômenos a partir delas, exigindo mudanças e fazendo atualizações.

Referências

BRAGA, José Luiz. Neomaterialismo & Antropológicas. *Revista Galáxia*, São Paulo, 2020, n. 45, set-dez, p.20-33. Disponível em: <https://www.scielo.br/j/gal/a/dhB9X7k7p9gq7xDCdtXQpVc/?lang=pt>. Acesso em: 30 jun. 2023.

CARRERA, Fernanda. Racismo e sexismo em bancos de imagens digitais: análise de resultados de busca e atribuição de relevância na dimensão financeira/profissional. In: SILVA, Tarcízio (org.). *Comunidades, Algoritmos e Ativismos Digitais: Olhares Afrodiaspóricos*. São Paulo: Editora LiteraRUA, 2021, ed. 2, p. 147-165. E-book.

CORRÊA, Laura Guimarães. BERNARDES, Mayra. “Quem tem um não tem nenhum”: solidão e sub-representação de pessoas negras na mídia brasileira. In: CORRÊA, Laura Guimarães (org.). *Voices Negras em Comunicação: Mídia, racismos, resistências*. Belo Horizonte: Autêntica, 2020. v.1, p.203-220.

CORRÊA, Laura Guimarães. Interseccionalidade: um desafio para os Estudos Culturais na década de 2020. In: SANTOS, L. H. S; KARNOPP, L. B; WORTMANN, M. L. *O que são estudos culturais hoje? Diferentes praticantes retomam a pergunta do International Journal of Cultural Studies*. São Paulo: Pimenta Cultural, 2022. p.123-141. Disponível em: <https://www.lume.ufgs.br/bitstream/handle/10183/253452/001159493.pdf?sequence=1>. Acesso em: 30 jun. 2023.

DANTAS, Ricardo. O que é Midjourney AI e como funciona? *Teg6*, 15 mai. 2023. Disponível em: <https://teg6.com/106991/noticias/o-que-e-midjourney-ai-e-como-funciona/>>. Acesso em: 8 jul. 2023.

GILLESPIE, Tarleton. A relevância dos algoritmos. *Revista Parágrafo*, 2018, v. 6, n. 1, p. 95-121. Disponível em: <http://revistaseletronicas.fiamfaam.br/index.php/recicofi/article/view/722/563>. Acesso em: 30 jun. 2023.

HALL, Stuart. *Cultura e Representação*. Rio de Janeiro: Ed. PUC-Rio: Apicuri, 2016. p. 139-223.

LEMONS, André; BITENCOURT, Elias. Sete pontos para compreender o neomaterialismo. *Revista Galáxia*, São Paulo, 2021, n. 46, p. 1-10. Disponível em: <https://revistas.pucsp.br/index.php/galaxia/article/view/52017>. Acesso em: 30 jun. 2023.

LEMONS, André. Epistemologia da comunicação, neomaterialismo e cultura digital. *Revista Galáxia*, São Paulo, 2020, n. 43, jan-abr, p. 54-66 Disponível em: <https://www.scielo.br/j/gal/a/DvNQBJKxf4hBZf3c-QHBL5FL/?lang=pt>. Acesso em: 30 jun. 2023.

MARQUES, Ana. VINHA, Felipe. *O que é Discord? Conheça funções, planos e dicas de segurança*. Tecnoblog, 2020. Disponível em: <https://tecnoblog.net/responde/o-que-e-discord/>. Acesso em: 8 jun. 2023.

MIDJOURNEY, programa por trás de imagens falsas que viralizaram, suspende testes gratuitos. *Revista Pequenas Empresas, Grandes Negócios (PEGN)*, 30 mar. 2023. Disponível em: <https://revistapegn.globo.com/tecnologia/noticia/2023/03/midjourney-suspende-testes-gratuitos-apos-imagens-falsas-viralizarem.gh.html>. Acesso em: 8 jun. 2023.

SALKOWITZ, Rob. *Midjourney Founder David Holz On The Impact Of AI On Art, Imagination And The Creative Economy*. *Forbes*, 16 sep. 2022. Disponível em: <https://www.forbes.com/sites/robsalkowitz/2022/09/16/midjourney-founder-david-holz-on-the-impact-of-ai-on-art-imagination-and-the-creative-economy/?sh=1011152d2d2b>. Acesso em: 8 jun. 2023.

SILVA, Tarcízio. *Racismo Algorítmico em Plataformas Digitais: microagressões e discriminação em código*. In: SILVA, Tarcízio (org.). *Comunidades, Algoritmos e Ativismos Digitais: Olhares Afrodiaspóricos*. São Paulo: Editora LiteraRUA, 2021, ed. 2, p 127-145. Disponível em: https://www.researchgate.net/publication/339954112_Comunidades_Algoritmos_e_Ativismos_Digitais_olhares_afrodiasporicos. Acesso em: 30 jun. 2023.

STANLEY-BECKER, Isaac. HARWELL, Drew. *Conheça a Midjourney, startup de IA por trás das imagens falsas do Papa Francisco e de Donald Trump*. *Byte*, 17 abr. 2023. Disponível em: <https://www.terra.com.br/byte/conheca-a-midjourney-startup-de-ia-por-tras-das-imagens-falsas-do-papa-francisco-e-de-donald-trump,07a5b975700c143874185b8d8c4a524edat74ecu.html>. Acesso em: 8 jun. 2023.



■ Giselle Beiguelman, frames do vídeo generativo da série Flora Rebellis, versão 2024.
Imagens geradas com Inteligência Artificial (Style Gans)

■ [Acesse o vídeo clicando na imagem](#)





O PENSAMENTO NÃO OCIDENTAL, A COSMOTÉCNICA DE YUK HUI E A TECNOLOGIA CHINESA

DOI: <https://doi.org/10.35699/2965-6931.2023.47714>

BEATRIZ CARMO E SILVA*; MARCELA VAROTTO MARAJÓ**; MARCELO REIS MAIA***; NATACHA RENA****

RESUMO: Os debates sobre os avanços tecnológicos têm levantado diversas questões ao longo dos últimos anos. No entanto, destaca-se a ideia de que a tecnologia não é universal, mas moldada por diferentes cosmotécnicas. Yuk Hui contesta a visão ocidental de tecnologia como uma expressão universal do ser humano, propondo que diferentes cosmovisões influenciam o desenvolvimento tecnológico. O autor explora a relação entre as cosmotécnicas e o pensamento chinês tradicional, incluindo o taoísmo e o confucionismo. O presente artigo evidencia o rápido desenvolvimento tecnológico na China, especialmente na era de Xi Jinping, responsável por tornar o país uma potência mundial que lidera pesquisas em várias frentes tecnológicas, incluindo Inteligência Artificial (IA). A partir desse contexto, busca-se compreender de que maneira o pensamento tradicional chinês se relaciona com a expansão tecnológica do país, principalmente com os projetos de Inteligência Artificial (IA), através das reflexões propostas por Yuk Hui.

PALAVRAS CHAVE: cosmotécnica; inteligência artificial; China. tecnologia; pensamento tradicional.

Non-western thinking, Yuk Hui's cosmotechinics and the chinese technology

ABSTRACT: Debates on technological advances have raised several issues over the past few years. However, the idea that technology is not universal, but shaped by different cosmotechinics, stands out. Yuk Hui challenges the Western view of technology as a universal expression of the human being, proposing that different worldviews influence technological development. The author explores the relationship between cosmotechinics and traditional Chinese thought, including Taoism and Confucianism. This article highlights the rapid technological development in China, especially in the era of Xi Jinping, responsible for turning the country into a world power, leading research on various technological fronts, including Artificial Intelligence (AI). From this context, we seek to understand how traditional Chinese thinking relates to the country's technological expansion, especially with Artificial Intelligence (AI) projects, through the reflections proposed by Yuk Hui.

KEYWORDS: cosmotechinics; artificial intelligence; China; technology; traditional thinking..

* Universidade Federal de Minas Gerais/ ** Universidade Federal de Minas Gerais/ *** Universidade Federal de Minas Gerais/

**** Universidade Federal de Minas Gerais

I A Revolução Chinesa foi antecedida por uma série de acontecimentos, como as Guerras do Ópio, o fim da monarquia e uma guerra civil, e um período de instabilidade. Para Mão Jr. (2010): “Os comunistas, fortalecidos pela sua brilhante atuação na II Guerra Mundial, rapidamente consolidaram o controle das áreas rurais, colocando a maioria das cidades chinesas sob cerco. Com o tempo este cerco foi se apertando e nas principais cidades chinesas começaram a ocorrer constantes agitações operárias e estudantes. Paulatinamente as cidades chinesas vão caindo nas mãos dos comunistas e Chiang Kai-Shek fugiu para a Ilha de Taiwan. Em Primeiro de Outubro de 1949 as tropas do Exército Popular de Libertação marcharam vitoriosas. Neste mesmo dia o camarada Mao Tsé-Tung proclamou a fundação da República Popular da China. Mais de um século transcorreu desde que as antigas estruturas do sistema imperial entraram em crise diante da penetração estrangeira. Muitas décadas se passaram desde que o Império ruiu e mergulhou a China no mais completo caos e guerra civil. Muitos Anos se passaram sem que a China tivesse um Governo que exercesse a soberania sobre todo o seu território. Pela primeira vez depois de muitas décadas, o Tiang Ming estava restabelecido. Mas diferentemente dos governos anteriores, não eram os mandarins imperiais ou títeres de potências

Introdução

Os avanços tecnológicos, há muito, suscitam discussões acerca do seu processo de desenvolvimento e de suas consequências. As pesquisas, no entanto, tendem à universalização no que diz respeito à caracterização desse processo em termos globais. Isso porque não se compreende a marcha tecnológica como um caminho múltiplo e não linear, muitas vezes distante do que o ocidente entende enquanto modernização. Nesse contexto, Yuk Hui traz o conceito de cosmotécnica. O autor apresenta uma antítese ao que se entende por tecnologia enquanto uma exteriorização antropológica do ser humano universal, que reside na ideia de que a tecnologia não é antropológicamente universal, mas que é atravessada por cosmologias particulares, para além da funcionalidade. Ou seja, existem várias cosmotécnicas no lugar de uma tecnologia única.

A cosmologia é um ramo da astronomia que estuda a estrutura do universo, desde sua origem e composição. A partir da ideia de cosmos enquanto compreensão acerca do universo, entende-se as cosmotécnicas como os diferentes universos tecnológicos que estão em desenvolvimento. Aqui destacam-se as realidades não ocidentais ou eurocentradas. Para Hui (2016):

Existe um equívoco geral de que todas as técnicas são iguais, de que todas as habilidades e produtos artificiais de todas as culturas podem ser reduzidos a uma coisa chamada “tecnologia”. E, de fato, é quase impossível negar que as técnicas podem ser entendidas como extensão do corpo ou externalização da memória. No entanto, eles podem não ser percebidos ou refletidos da mesma forma em diferentes culturas. (Hui, 2016, p. 9, tradução nossa).

Yuk Hui também aborda o taoísmo e o confucionismo, duas das principais correntes filosóficas chinesas, como pontos de partida para a compreensão da evolução tecnológica, principalmente na China. Para o pensamento tradicional chinês, existem duas categorias principais: “*dao*” (ou *tao*), a força vital que permeia todas as coisas, também compreendida como caminho; e “*qi*” (ou *chi*), que significa utensílio ou ferramenta que tem a capacidade de promover uma harmonia com a natureza. O autor defende, no entanto, que a tecnologia ocidental vem sendo desenvolvida como ferramenta de sobreposição à natureza. Nesse sentido, afirma:

Ao longo da história chinesa, a unidade compreendida de *dao* e *qi* constituiu a moralidade e a forma de vida própria de cada época sucessiva. Essa unidade motivou e restringiu o desenvolvimento da tecnologia na China em comparação com o Ocidente, onde a tecnologia foi impulsionada pela razão instrumental por meio da qual as ferramentas são moldadas como um meio de superar em vez de se harmonizar com a natureza. (Hui, 2020, tradução nossa).

Se nos voltarmos ao oriente como campo de estudo do desenvolvimento das tecnologias mais avançadas num movimento de compreender esse fenômeno como “multi-cósmico”, é quase impossível, e até injusto, não ter foco sobre a China. Isso porque nenhum país passou por transformações tão profundas num dado período de tempo quanto tal país.

A história recente da China inicia-se na década de 1940 com a Revolução Chinesa¹ sob a liderança de Mao Tsé-Tung. A ascensão do Partido Comunista Chinês (PCCh) ao poder deu início a uma série de mudanças estruturais da sociedade chinesa que levaram o país à posição de segunda maior economia mundial atualmente. Nesse contexto, foi Xi Jinping² que ascendeu como principal presidente e líder do acelerado crescimento do país. A economia da China mais que dobrou sob a gestão de Xi, indo de US\$ 8,53 trilhões em 2012, seu primeiro ano como presidente, para US\$ 17,73 trilhões em 2021. A tecnologia também se tornou um campo essencial do desenvolvimento chinês, havendo planos para a liderança mundial da China no campo da Inteligência Artificial (IA), como está ocorrendo nos últimos anos.

Segundo o *Critical Technology Tracker* publicado pelo Australian Strategic Policy Institute (ASPI)³, há uma impressionante liderança chinesa em pesquisas de alto impacto e está a frente em 19 das 23 tecnologias prioritárias definidas pela AUKUS⁴, parceria de defesa entre Austrália, Grã-Bretanha e Estados Unidos. Incluem-se nas supracitadas prioridades os drones submarinos e hipersônicos (Needham, 2023). Como apresentado por Laura Dobberstein (2023), a liderança desempenhada pela China é um produto direto dos projetos e das políticas do PCCh, especialmente os de longo prazo, promovidos por Xi Jinping e outros importantes líderes do país e do partido.

Como citado acima, um dos campos mais recentes quando se trata dos avanços tecnológicos é o da Inteligência Artificial (IA). De forma resumida, pode-se assumir

estrangeiras que governavam: agora quem estava no poder era o povo.” (Máo Jr., 2010, p. 6).

2 Xi Jinping é o presidente da China desde 2013 e Presidente da Comissão Militar Central desde 2012, sendo o líder político chinês mais proeminente dos últimos anos. Nascido em Beijing no ano de 1953, Xi completou 70 anos em 15 de junho de 2023. Sua trajetória política se iniciou precocemente como secretário no serviço militar nos anos 1979, com 26 anos. Graduou-se em 2002 com doutorado em direito e ideologia, tendo estudado teoria marxista e educação ideológica na Universidade de Tsinghua. Em 2007, Xi foi nomeado para o Comitê Permanente composto por nove membros no 17º Congresso do Partido. Desde então, Xi viu sua participação política crescer e passar para vice-presidente no ano de 2009 e, finalmente, para presidente em 2013.

3 O Australian Strategic Policy Institute (em português, Instituto Australiano de Política Estratégica) é considerado um think tank e é financiado pelo governo australiano desde sua criação em 2001.

4 A AUKUS é uma aliança militar firmada entre a Austrália, os EUA e o Reino Unido. O principal objetivo é conter o expansionismo chinês na região Indo-Pacífica.

os esforços da IA como a automatização extrema das máquinas, que nos leva ao termo “aprendizado de máquina”. A partir dos estudos da Inteligência Artificial, as máquinas evoluem de meras ferramentas para instâncias que possuem uma espécie de raciocínio e se aproximam cada vez mais da dinâmica de funcionamento do cérebro humano. Nesse sentido, as redes neurais exemplificam a operação da IA. Isso porque elas foram inspiradas no funcionamento dos neurônios humanos e permitem o aprendizado profundo capaz de desenvolver técnicas computacionais semelhantes às camadas neurais.

Propõe-se nesse artigo, portanto, que se faça uma análise dos projetos chineses envolvendo Inteligência Artificial a fim de compreender de que maneira o pensamento tradicional chinês, influenciado pelas correntes taoísta e confucionista, contribui para o desenvolvimento tecnológico do país. Apoiados na filosofia de Yuk Hui (2016; 2020; 2021), buscamos elucidar como a tecnologia não tem uma compreensão universal e pode se dar de maneiras diferentes daquelas conhecidas pelo ocidente. O filósofo nos convida a expandir nossa capacidade cognitiva ao aprender com outras cosmotécnicas

Inteligência chinesa e pensamento não ocidental *Yuk hui e as considerações sobre modernidade*

Ronaldo Lemos, no prefácio da versão em língua portuguesa do livro *Tecnodiversidade* de Yuk Hui (2020), afirma que a decisão de ler ou não a obra do autor é como uma escolha entre duas pílulas: onde a primeira, de não ler, o mantém na confortável ignorância, como uma “prisão cognitiva”, enxergando a tecnologia como força única que avança por si só; a segunda, que significa a leitura, por sua vez, é a pílula que aprofunda a reflexão acerca da tecnologia e que delineia uma realidade mais dura, de fato, mas libertadora. Aqui se objetiva a abertura da discussão para compreender o desenvolvimento tecnológico que foge da pré-concepção ocidental e se volta para outras formas de saberes.

Para a reflexão proposta a partir de Yuk Hui, é necessário, antes de tudo, compreender a discussão que o autor traz à tona acerca da modernidade e da decadência da ideia de cosmos:

A “morte do cosmos” de Brague, empreendida pela modernidade europeia e por

sua globalização da tecnologia moderna, necessariamente forma uma das condições para que hoje reflitamos sobre a cosmopolítica, uma vez que ilustra a ineficácia de uma metáfora biológica para o cosmopolitismo. (Hui, 2020, p. 21).

Diante da “morte do cosmos” atrelada à globalização universalizante que está embebida na ideia de modernidade eurocentrada, o autor sugere a “virada ontológica” que se constitui como uma tentativa de dar voz a diferentes ontologias⁵. Para Hui, a virada ontológica é liderada, entre outros autores, por “Philippe Descola, Eduardo Viveiros de Castro, Bruno Latour e Tim Ingold, e, antes deles, Roy Wagner e Marilyn Strathern.” (Hui, 2020, p. 21).

Essa virada ontológica é uma resposta direta à crise da modernidade que, de modo geral, se expressa em termos de uma crise ecológica que, agora, está intimamente ligada ao Antropoceno. O movimento da virada ontológica é uma tentativa de levar diferentes ontologias em diferentes culturas a sério (devemos ter em mente que saber onde diferentes ontologias estão não é o mesmo que levá-las a sério). (Hui, 2020, p. 21-22).

O que Hui elucida a partir da supracitada virada ontológica é uma superação das dicotomias muito presentes no pensamento majoritariamente europeu considerado moderno. Para além de um pensamento pós-colonial, o autor sugere a concretização de um pluralismo ontológico a partir de questões relacionadas à tecnologia e à política. Os esforços do presente trabalho vão ao encontro desse posicionamento e buscam dar luz aos projetos e concepção de tecnologias a partir de uma outra visão.

A antítese é onde podemos identificar e posicionar o não-moderno. É aquilo que se recusa a ser reduzido a um processo histórico linear e que resiste à expulsão do mundo. No primeiro volume de *Science and Civilisation in China*, o grande sinólogo Joseph Needham afirmou várias vezes que não é produtivo comparar a ciência e a tecnologia chinesas com as do Ocidente como se fossem iguais, uma mais avançada do que a outra em diferentes períodos de uma história linear e única. No segundo volume, com o subtítulo *History of Scientific Thought*, Needham continua a sua afirmação de que o pensamento científico e tecnológico chinês é fundamentalmente organístico e que, ao contrário dos primeiros modernos europeus, por exemplo, Descartes, nunca foi mecanicista. (Hui, 2021, tradução nossa).

Hui (2020) em seu livro *Tecnodiversidade* busca construir uma nova “genealogia do pensamento tecnológico chinês” (Hui, 2020, p. 27) e sugere que todos os países

⁵ Entende-se por ontologia o ramo da filosofia que se dedica ao estudo sobre o ser, sua existência, o devir e a realidade. Ou seja, quando se propõe uma coexistência de diferentes ontologias, refere-se às questões que circundam o mundo e sua compreensão pelos seres humanos. Para Antonio Henrique Bernardes (2022): “[...] tratamos a Metafísica e a Ontologia como tradições filosóficas próximas e constituintes de um mesmo movimento que possui como base as reflexões sobre os modos de existência dos homens.” (Bernardes, 2022, p. 428).

também o façam. O que o autor traz acerca da cosmotécnica chinesa é a união de duas instâncias que remontam ao pensamento tradicional chinês, muito relacionado ao taoísmo e ao confucionismo: o qi (ou chi) e o dao (ou tao). Para ele:

A união do qi e do dao também é a união da moral e do cósmico, já que a metafísica chinesa é, em essência, uma cosmologia moral ou uma metafísica moral, como foi demonstrado pelo filósofo do novo confucionismo Mou TsungSan. Mou sugere que, se em Kant podemos encontrar uma metafísica da moral, se trata no máximo de uma exploração metafísica da moral, não uma metafísica moral, uma vez que esta só pode ter início com a moral. (Hui, 2020, p. 27).

A descoberta da cosmotécnica está relacionada a um contexto que vai além da própria técnica e que se une a outros aspectos de uma sociedade ou uma nação, a saber: cultura, política, costumes e diversos outros âmbitos que não podem ser explicados sem a plena vivência.

A ideia a ser superada é a de que o avanço tecnológico se dá dentro de um processo linear e cumulativo e que, por esse motivo, existem sociedades mais ou menos civilizadas quando analisadas por esse viés. Além disso, busca-se o afastamento da noção de que a tecnologia avançada pressupõe o domínio em relação à natureza, criando uma hierarquia entre cultura e natureza. Do mesmo modo que o giro ontológico, utilizado por Hui, evocava uma compreensão ampla acerca da natureza, a cosmotécnica, por sua vez, chama para si uma noção múltipla de tecnologia e, conseqüentemente, se volta ao solo oriental.

O maior exemplo utilizado para a explicação acerca do conceito de cosmotécnica é a história do açougueiro Pao Ding, ou simplesmente cozinheiro Ding. No conto escrito por Zhuangzi, o profissional é conhecido pelos seus talentos de cortar e desmembrar animais bovinos sem tocar seus tendões e ossos. Quando há o questionamento acerca da sua habilidade, o açougueiro responde que o que ele ama é o dao, muito mais esplêndido que a técnica (The Dewdrop, 2020). Ou seja, o segredo da habilidade não está diretamente relacionada às ferramentas que ele utiliza, mas que existe algo muito além do que a instrumentalização pode oferecer e que flui naturalmente pelas mãos do cozinheiro. Portanto, “a razão instrumental, que poderia se entender casualmente como a lógica que unifica os movimentos individuais com resultados individuais, parece fora de jogo.” (Instituto Humanitas Unisinos, 2020).

O sinofuturismo também é um conceito importante dentro do presente contexto.

O termo remete à ideia de um futuro majoritariamente representado pelas técnicas, tecnologias e ferramentas chinesas. Porém, cabe ressaltar que, para Hui, o sinofuturismo não parece muito frutífero, uma vez que apenas reproduz a lógica capitalista ocidental de corrida pela liderança tecnológica. O que importa, de fato, é a cosmotécnica chinesa por si só e não o que ela representa em termos de avanços tecnológicos a partir da noção ocidental europeia.

Enquanto Yuk Hui desenvolve sua obra bastante conectado à filosofia, pretende-se, no presente trabalho, construir uma investigação que possa, através de exemplos concretos de projetos de desenvolvimento tecnológico ancorados em Inteligência Artificial na China, demonstrar o funcionamento dessa outra cosmotécnica. Para isso, considera-se necessário compreender melhor os conceitos do pensamento tradicional chinês, principalmente do taoísmo e do confucionismo, muito utilizados por Hui no desenvolvimento do seu trabalho.

Outros saberes: taoísmo e confucionismo

O taoísmo e o confucionismo são saberes tradicionais chineses muito usados por Yuk Hui em suas reflexões acerca da sociedade tecnológica. Não se pretende aqui apresentar uma definição estrita do taoísmo ou do confucionismo, mas introduzir aspectos relevantes da crença histórica que inspiram as reflexões de Hui e que podem ser férteis para a análise proposta no presente trabalho.

O que é muito utilizado pelo autor e diz respeito às supracitadas correntes são, principalmente, o dao (ou tao) e o qi (ou chi). Pode-se dizer que o dao é a força motriz por trás de todas as coisas, aquilo que dá sentido, que constrói o caminho. Em relação ao qi, tem-se um elemento básico de tudo que existe e representa o poder funcional dos eventos. Portanto, é possível compreender o dao enquanto um tipo de essência, mas que não pode ser controlado, que flui naturalmente e está acima de todas as decisões; já o qi relaciona-se ao que se aproxima da matéria, da ferramenta (Figura 1).

Portanto, as ferramentas não são pensadas como algo desapegado, completamente autônomo, mas sim como recipientes, como containers. É assim que *qi* necessita do *dao*, e vice-versa. “*Qi*”, às vezes também se traduz como “coisas materiais”, “o que está debaixo da forma”. Os utensílios, entendidos como recipientes, requerem então, quase

por definição, algo “mais além da forma” que funcione como seu conteúdo. (Instituto Humanitas Unisinos, 2020).

Um grande exemplo a ser utilizado ao se tratar da influência do pensamento tradicional chinês, especificamente do taoísmo, no desenvolvimento da sociedade é a medicina. Enquanto a medicina moderna ocidental cura o corpo a partir da aplicação da ciência sobre o corpo de forma mecânica, a medicina tradicional chinesa tem por cura o alinhamento e harmonia dentro do corpo. Para Hui (2020):

A medicina chinesa é baseada na cosmologia taoísta, com Yin, Yang e cinco tipos de Qi. Se, de uma perspectiva ocidental, você abordar um médico chinês e perguntar: “Você pode me mostrar seu Qi e provar que essa energia existe?”, a resposta teria que ser não. Se você não pode provar a existência da energia que está na base de sua prática, como pode dizer que pratica uma ciência? Aqui reside o problema. Mas isso não significa que a medicina chinesa não seja científica. Como ciência empírica, funcionou por 2.000 anos com base em uma epistemologia diferente. Por muito tempo, em Hong Kong, a medicina chinesa foi classificada abaixo da medicina ocidental. Se você for a um médico chinês, isso não será coberto pelo seu seguro de saúde porque a medicina chinesa é vista como não científica. (Hui, 2020, tradução nossa).

O pensamento tradicional chinês na reflexão proposta por Yuk Hui é um convite a um caminho oposto ao da universalização que surge da globalização, essa que, por sua vez, foi demasiadamente impactada pelo avanço tecnológico dos últimos anos. Pode-se afirmar que a intenção é, na verdade, se voltar a outros aspectos de uma sociedade considerada tecnologicamente avançada — dentro da hierarquia que se cria quando considerada a partir de um processo linear exponencial — e compreendê-la dentro de um complexo e multifacetado processo.

A ética do confucionismo garantia a legitimidade do poder imperial. Segundo esta ética, o Imperador era detentor do Mandato Celestial (Tiang Ming). Entretanto, se o Imperador se mostrasse inepto, os Céus lhe retirariam o poder, ocorrendo então a Ruptura do Mandato Celestial (Ge Ming). Segundo a lógica confuciana o Ge Ming era prenunciado por “sinais”, tais como prodígios no céu, catástrofes climáticas, corrupção na burocracia e agitação agrária. Destes “sinais” devemos nos ater ao que nos parece mais importante: a agitação agrária. (Máo Jr., 2010, p. 4).

Além disso, o pensamento tradicional chinês, principalmente no que diz res-

peito ao confucionismo, tem enorme influência em toda a história da cultura do país. Durante o período próximo à Revolução Chinesa, no entanto, esse pensamento perdeu o crédito que possuía até então, fato relacionado a um diagnóstico de que a tradição poderia ser danosa aos projetos de modernização chinesa.

A Revitalização de Confúcio de Xi faz parte de uma tentativa mais ampla de reintrodução da cultura tradicional chinesa. Como parte de um longo ciclo de “neotradicionalismo”, Xi está tentando reviver antigas culturas e valores para uso político moderno. A cultura tradicional chinesa tem visto um renascimento desde a virada do século, mas sob Xi isso foi ampliado. Enquanto Mao via a cultura tradicional como a “velha sociedade”, Xi a vê como a “alma” da nação. O impulso neotradicionalista de Xi faz parte de seu “Sonho da China”, que vê o “grande rejuvenescimento da nação chinesa”. Xi busca devolver a China a um lugar de glória nacional, com objetivos territoriais e econômicos. (Clayton, 2020, tradução nossa).

O confucionismo é uma religião nomeada em homenagem ao seu criador: Confúcio. O filósofo, que viveu na China entre os anos 551 e 479 a.C., não recebeu grande reconhecimento em vida, tendo suas ideias perpetuadas algum tempo depois e amplamente estruturadas e enraizadas na cultura chinesa. O confucionismo também pode ser considerado uma filosofia de vida ou um sistema ético e, por esse motivo, dispõe de alguns ideais principais: a família, o mundano, a tradição e o ritual, muito ligados à busca do bem-estar comum e da criação de uma sociedade culturalmente instruída (Sciulo, 2020).

Entende-se, portanto, que o pensamento tradicional chinês passou por diferentes momentos ao longo da história, mas nunca deixou de ser um ponto importante de análise da cultura e do modo de vida chineses. No contexto do presente trabalho, chama-se a atenção para o interesse de Xi Jinping de retomar a tradição e, de alguma forma, incorporá-la no seu modo de governo. O que, consequentemente, tem influência nas diversas áreas que uma gestão nacional abarca, incluindo os projetos de inovação e tecnologia.



Figura 1 - Ideogramas de Qi e Dao respectivamente. Fonte: Wikipedia.

O desenvolvimento tecnológico chinês

Numa tentativa de compreender o contexto de desenvolvimento tecnológico chinês à maneira de sua própria cosmotécnica, é imprescindível que se atenha ao processo que precedeu esse tipo de avanço e que colocou o país na liderança mundial dos investimentos em tecnologia. Os projetos chineses, materializados como robôs, sensores e dispositivos, são a ferramenta, ou seja, o seu *qi* e, como *qi* não se traduz de forma individual, busca-se compreender o *dao* nessa conjuntura cosmotécnica, o que, de fato, se mostra como caminho e está além da própria materialidade.

Tentei entender a cosmotécnica chinesa [tecnologia que implica uma visão de mundo] por meio da relação dinâmica entre duas grandes categorias do pensamento tradicional chinês. “Dao”, ou a força vital etérea que circula todas as coisas (comumente chamada de caminho), e “Qi”, que significa ferramenta ou utensílio. Juntos, dao e qi — a alma e a máquina, por assim dizer — constituem uma unidade inseparável. Ao longo da história chinesa, a unidade compreendida de dao e qi constituiu a moralidade e a forma de vida própria de cada época sucessiva. Essa unidade motivou e restringiu o desenvolvimento da tecnologia na China em comparação com o Ocidente. Lá, a tecnologia foi impulsionada pela razão instrumental, por meio da qual as ferramentas são moldadas como um meio de superar, em vez de se harmonizar com a natureza. (Hui, 2020, tradução nossa).

A conjuntura cosmotécnica chinesa, na qual se deu o desenvolvimento tecnológico chinês, tem gerado incontáveis produtos com tecnologias inovadoras, surpreendendo o mercado a cada ano. As suas tecnologias de reconhecimento de voz e reconhecimento facial, por exemplo, são líderes mundiais. No entanto, as inovações, tanto digitais como eletrônicas, não tiveram como ponto de partida as iniciativas individuais e isoladas das Big Techs chinesas que comercializam as tecnologias como produtos. A China, desde 2017, sinaliza em seus relatórios governamentais que a Inteligência Artificial é uma estratégia nacional do país e, portanto, um objetivo a ser perseguido por toda a nação.

No âmbito dos objetivos estratégicos da China, especialmente dos que dialogam com a expansão do desenvolvimento tecnológico, evidencia-se a atuação de empresas de tecnologia no território chinês. Aqui, destacam-se: Baidu, Alibaba e Tencent (conhecidas como BAT) e também, iFlytek. As Big Techs chinesas estão continuamente em processo de pesquisa e desenvolvimento para o lançamento de plataformas apoiadas em Inteligência Artificial. Pontua-se que, para o governo chinês, a IA é uma das alternativas para reconstruir os elos das atividades econômicas, como produção, distribuição, troca e consumo, haja visto que ela se tornou o novo motor do desenvolvimento econômico no mundo.

O trabalho conjunto do Governo e das Big Techs ganhou força após o discurso de abertura do 19º Congresso Nacional do Partido Comunista da China (PCCh). O presidente chinês, Xi Jinping, afirmou que a inovação e a tecnologia são o meio de superar o gargalo no desenvolvimento econômico da China (China, 2016). Nesse sentido, as gigantes de tecnologia deram início a uma complexa operação de desenvolvimento de tecnologias avançadas, incluindo a Internet, o Big Data e a IA, a fim de alavancar a economia chinesa. Isso demonstrou que tais empresas trabalham em concordância com os objetivos do Estado.

Um dos princípios básicos para o desenvolvimento tecnológico chinês é o código aberto. Essa iniciativa permite promover a co-criação de tecnologias e o compartilhamento de vários assuntos de inovação entre a indústria e a academia através da produção científica (Gráfico 1). Contudo, nesse processo colaborativo, os dados coletados por pesquisas, isto é, qualquer registro de informação em meio eletrônico ou em qualquer outra forma são protegidos continuamente pelo Estado chinês. Esse procedimento é

nomeado «segurança de dados» e também se aplica para o processamento dos dados, o qual inclui o uso, a transmissão, o fornecimento e a divulgação. A Lei de Segurança de Dados da República Popular da China, promulgada em 2021, estabelece tais diretrizes a fim de salvaguardar a soberania, a segurança e os interesses de desenvolvimento do país. Aqui percebemos claramente que o Dao – o caminho que a tecnologia percorre – diverge do Ocidente onde os dados não estão sob a proteção de nenhum Estado ocidental.

Nota-se também que o um planejamento do Estado chinês, fomenta a atuação conjunta de instituições e empresas na elaboração de uma solução tecnológica contribui com o desenvolvimento econômico coordenado e direcionado, uma vez que o surgimento de novas tecnologias altamente eficientes aquece e movimenta o mercado. Assim, nota-se uma transformação da sociedade resultante da aplicação de conhecimentos científicos de múltiplos campos em áreas estratégicas.

É importante mencionar que a nova demanda por inteligência em vários campos, do macro ao micro, foi um dos fatores que impulsionou a criação de novas tecnologias, novos produtos, novas indústrias e novos formatos. Como consequência disso, grandes mudanças na estrutura econômica foram desencadeadas de modo a alterar profundamente a produção humana, os estilos de vida e os padrões de pensamento, além de ter promovido um salto geral na produtividade social.

No documento intitulado “Plano de Desenvolvimento de Inteligência Artificial de Nova Geração” (New Generation Artificial Intelligence Development Plan - NDAIDP), emitido pelo Conselho Estadual chinês em 2017, é elucidado que a IA é uma ferramenta importante para a manutenção da segurança nacional e que o desenvolvimento da Inteligência Artificial contribuirá para construir o pioneirismo do país, acelerando a construção de um país inovador e de uma potência mundial em ciência e tecnologia. O Conselho Estadual sinaliza que alguns dos objetivos para o futuro da China são:

[...] colocar o desenvolvimento da Inteligência Artificial no nível estratégico nacional, planejando-a sistematicamente e ativamente, participar ativamente da pesquisa e do desenvolvimento global de uma governança de inteligência artificial e otimizar a alocação de recursos de inovação em escala global. (China, 2017, tradução nossa).

O governo chinês indica a Inteligência Artificial como uma estratégia nacional e elenca etapas para que isso se efetive. Primeiramente, o país estabeleceu como meta a equiparação das tecnologias, incluindo a IA, aos níveis mais avançados existentes até o ano de 2020, a fim de que a indústria da Inteligência Artificial se tornasse um dos pilares do seu crescimento econômico da China e de que essa tecnologia fosse capaz de melhorar a vida das pessoas ao transformar seus meios de subsistência. Isso permitirá que o país esteja entre os *rankings* e as listagens dos países detentores de tecnologias inovadoras e possua uma sociedade próspera de maneira abrangente.

Em segundo lugar, é apresentado entre os objetivos estratégicos da China o avanço na teoria básica da Inteligência Artificial até 2025 e o desenvolvimento de tecnologias e soluções em níveis de liderança mundial. Com isso, a IA se tornará a principal força motriz para a modernização industrial e para a transformação econômica do país, acarretando progresso na construção de uma sociedade inteligente. Isso implica em assumir que a mão de obra precisa ser desenvolvida de modo coordenado. A China assume a substituição da mão de obra humana por IA e alimenta seu processo coordenado de desenvolvimento com soluções para este desafio. No ocidente, o processo de desenvolvimento tecnológico não coordenado deixa o trabalhador desamparado.

A terceira etapa, datada para 2030, consiste na conquista da liderança mundial por parte da China da teoria, tecnologia e aplicação da Inteligência Artificial de modo que o país se torne o principal centro de inovação em IA do mundo. Consequentemente, a economia inteligente e a sociedade inteligente — originadas a partir das transformações proporcionadas por esse recurso — terão maior notabilidade e irão servir como embasamento para que a China seja intitulada potência econômica e inovadora. Contudo, apesar do governo chinês ter estabelecido um prazo distante para que o país assumisse uma posição de destaque global na tecnologia de IA, observando a partir de uma perspectiva acadêmica, a China já se tornou o líder mundial em publicações e patentes envolvendo essa tecnologia.

6 Para acessar outros comparativos e rankings acerca de algoritmos de Inteligência ver <https://techtracker.aspi.org.au/tech/artificial-intelligence-algorithms-and-hardware-accelerators/?colours=true>.

7 Os Planos Quinquenais são um modelo de política da China para o desenvolvimento econômico e social de médio prazo.

8 O desenvolvimento de alta qualidade se efetiva por meio do «desenvolvimento de empresas privadas, da atualização e transformação industrial, da economia digital e da alta tecnologia» (CHINA, 2023).

9 O “socialismo de mercado” é uma formação que associa modos de produção de diferentes épocas históricas em clara unidade de contrários. É sinônimo do que se denomina “etapa primária do socialismo” (Jabbour; Dantas, 2020; Jabbour et al., 2017).

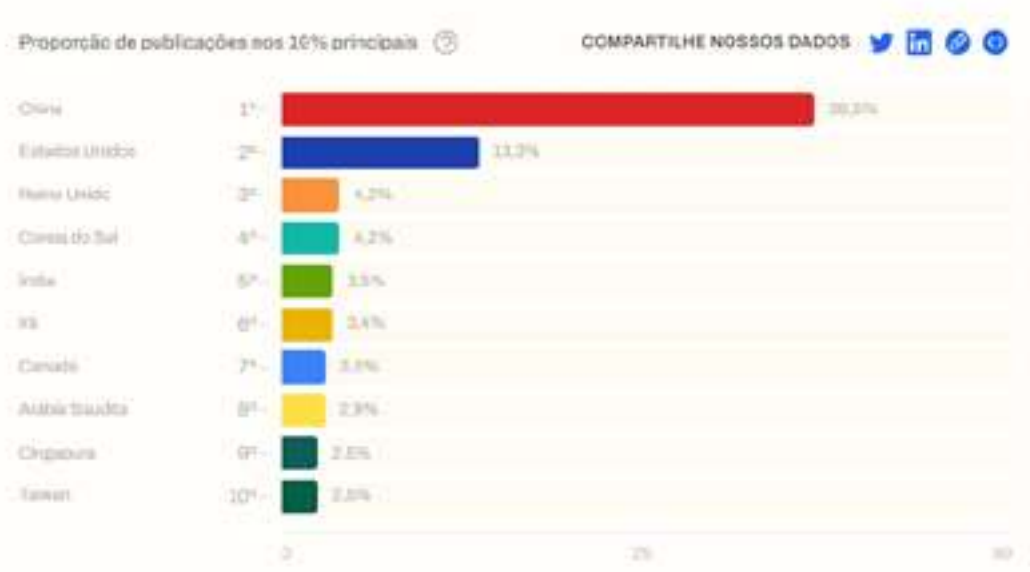


Figura 2 – Proporção de publicações por país nos 10% principais artigos citados sobre algoritmos de Inteligência Artificial. Fonte: ASPI (2023)⁶

Após a divulgação do NDAIDP, foi criado o Comitê Consultivo Estratégico de Inteligência Artificial de Nova Geração (New Generation Artificial Intelligence Strategic Advisory Committee). O grupo é responsável pela liderança no planejamento e na coordenação abrangentes do processo de desenvolvimento da IA, deliberando sobre as principais tarefas, políticas, questões e os principais arranjos de trabalho. Destaca-se que o comitê atua continuamente em parceria com o Ministério da Ciência e Tecnologia da China, outro importante ator no desenvolvimento tecnológico chinês. O trabalho conjunto da rede de atores apresentada permite, gradualmente, construir think tanks de Inteligência Artificial.

O Governo da China, nos anos seguintes a 2017, deu continuidade à elaboração de políticas que reforçassem a importância do desenvolvimento tecnológico chinês. Em 2021, foi apresentado o 14º Plano Quinquenal⁶ (14th Five-Year Plan), política que estrutura objetivos estratégicos nacionais a serem perseguidos num período de cinco anos — nesse caso, entre 2021 e 2025. O recorte temporal proposto se apresenta como uma importante oportunidade para inovação e liderança no desenvolvimento de alta qualidade⁷ (em inglês, high-quality development), conceito que vem sendo citado e enfatizado pelo governo chinês. Essa conjuntura envolve a construção de uma China

digital, resultante do desenvolvimento de uma economia centrada em produtos e serviços baseados no ambiente online.

O 14º Plano Quinquenal da China menciona a necessidade do fortalecimento do poder estratégico nacional⁸ por meio da ciência e da tecnologia com o apoio de laboratórios nacionais. Destaca-se que a iniciativa, por adotar uma escala nacional, implica o alinhamento e a colaboração entre universidades, institutos e centros de pesquisa, organizações e empresas, o que já previa o plano:

Promover o compartilhamento de recursos de institutos de pesquisa científica, instituições de ensino superior e forças corporativas de pesquisa científica. [...] Construir uma plataforma de intercâmbio de alto nível para trabalhos de pesquisa científica nacional e informações científicas e tecnológicas. (China, 2021).

Além disso, outros objetivos propostos na política citada são: melhorar as condições nacionais da economia socialista de mercado⁹, travar a batalha por tecnologias essenciais, melhorar a eficiência geral da cadeia de inovação, visar campos de ponta (como inteligência artificial, mecânica quântica, neurociência, reprodução biológica, tecnologia aeroespacial e exploração terrestre e marítima) e implementar uma série de projetos científicos e tecnológicos nacionais estratégicos voltados para o futuro (CHINA, 2021). Todas as metas dispostas consideram a atuação conjunta do Estado, por meio de empresas estatais, universidades e ministérios, e de empresas privadas em prol de avanços e ganhos voltados à nação.

Nesse sentido, é importante salientar o papel de planejador desempenhado pelo Estado chinês que está atrelado aos avanços tecnológicos da China ao longo do século XXI, emergidos em sua maioria após o NDAIDP e o 14º Plano Quinquenal. As iniciativas de planejamento perpassam por diversos âmbitos da sociedade como a economia, a ciência e a educação, a defesa nacional, a indústria, a tecnologia, entre outros. A significativa capacidade do Estado em intervir na realidade é um dos fatores, segundo Jabbour et al. (2022), que caracteriza o país como berço de uma nova formação econômico-social¹⁰. É importante salientar que a intervenção chinesa é planejada com base em projetos elaborados por equipes nacionais multidisciplinares, a fim de que demandas de diversas esferas possam ser contempladas.

[...] a predominância do planejamento estatal sobre os gânglios vitais da grande produção e da grande finança ao lado de uma soberania monetária particular permite ao Estado gerir um processo que entrelaça tanto uma maior restrição à ação da lei do valor quanto a transição de uma planificação orientada à geração de valor e ao mercado para o que chamamos de planejamento baseado no projeto. (Jabbour et al., 2022, p. 95).

Quando se discute sobre o modo de planejamento chinês contemporâneo, é imprescindível voltar a atenção ao *Projetamento*. Compreendemos o *Projetamento* como uma capacidade de planejamento exercida pelo Estado em um nível mais elevado, envolvendo previsões e a orientação de políticas e investimentos para estimular o progresso. Este conceito se refere ao nível superior, uma vez que transcende os planos e o planejamento convencional. A palavra «*projetamento*» (anteriormente conhecida como «*projectment*» em inglês arcaico) é mencionada no livro «Elementos de economia do projetamento» (1959) de Ignácio Rangel.

O *Projetamento* pode ser interpretado de diversas maneiras. Primeiramente, engloba o processo histórico no qual a prática de planejar grandes investimentos se tornou central nas economias capitalistas e socialistas. Nessas sociedades, economistas e engenheiros de projetos desempenham papéis cruciais no processo de intervenção estatal na economia. A financeirização e o colapso da União Soviética marcaram o fim das experiências das antigas economias baseadas no *projetamento*. Na China, o conceito ressurgiu vigorosamente por meio de uma economia na qual as forças da lei do valor são restringidas, abrindo espaço para uma economia centrada em grandes projetos.

Essa abordagem econômica também é caracterizada por sua ênfase na criação de bens públicos significativos (como o setor considerado improdutivo ou os «fundos de consumo» segundo Marx), bem como a construção de uma esfera voltada para o aprimoramento espiritual humano por meio de investimentos em saúde, educação, esporte, lazer, etc. No entanto, a noção de «projeto» pode ser compreendida apenas dentro dos limites de uma complexa operação contábil.

Partimos do pressuposto de que o socialismo ainda não alcançou um estágio de maturidade como modo de produção. Nesse contexto, o conceito de projeto e o *Projetamento* assumem uma essência caracterizada pelo uso da razão como instrumento governamental, representando a forma mais avançada historicamente pela qual o socialismo se apresenta ao mundo. O *Projetamento* (baseado na razão e no «governo fundamentado na ciência») se contrapõe à irracionalidade capitalista e à fetichização

do homem por si mesmo. Afinal, apenas o que não está sob controle se torna fetichizado. O Projetoamento representa um passo crucial na emancipação humana de suas próprias adversidades (Jabbour; Gabriele, 2021).

Contudo, ressalta-se que a China é marcada por uma nova leitura da economia do projetoamento: a Nova Economia do Projetoamento.

O que acontece na China é um estágio superior do domínio humano sobre a natureza (planejamento) e sobre a anarquia que é inerente às economias sob orientação capitalista. Esse nível superior de planejamento e toda a sua estrutura econômica e institucional é o que chamamos de «Nova Economia do Projetoamento», causa e consequência do papel central do Sistema Nacional de Inovações Tecnológicas da China desde a segunda metade do século XX. [...] Ao atingir as fronteiras tecnológicas em alguns aspectos críticos e socializar seu uso para o planejamento econômico em larga escala, esse sistema tornou-se o núcleo fundador da Nova Economia do Projetoamento. (Jabbour; Moreira, 2023, p. 546, tradução nossa.)

No âmbito da Nova Economia do Projetoamento, merecem destaque a planificação e a intervenção relacionadas à ciência e à tecnologia. O Estado chinês tem atuado fortemente no fomento à pesquisa e no desenvolvimento de tecnologias inovadoras (como a Inteligência Artificial, o *Big Data* e o 5G), o que já vem apresentando resultados como avanços no campo teórico a partir de descobertas científicas e criação de soluções tecnológicas apresentadas como produtos comercializáveis. Acerca disso, interessam-nos os projetos chineses desenvolvidos a partir da Inteligência Artificial (IA).

Projetos tecnológicos chineses

A Inteligência Artificial avançou em uma velocidade avassaladora nos últimos anos e tem influenciado o cotidiano das pessoas cada vez mais. Os *softwares* que utilizam IA são as ferramentas mais notáveis quando se trata do assunto, principalmente dentro de uma discussão eventualmente inexperiente acerca das tecnologias atuais. Não há dúvidas de que essas ferramentas são importantes e devem ser consideradas experiências bem sucedidas, ainda que careçam de amadurecimento acerca de sua utilização e regulamentação. Interessa-nos, todavia, a investigação a respeito dos projetos tecnológicos que objetivem mudanças na qualidade de vida. Nesse contexto, a partir do

que se expôs anteriormente, explora-se mais a fundo os projetos chineses de infraestrutura que utilizam Inteligência Artificial.

Primeiramente, convém compreender o conceito de infraestrutura. Definida pela junção entre o aparato físico e os serviços a ele associados presentes em uma determinada localidade, a infraestrutura contempla elementos fundamentais para um desenvolvimento socioeconômico. Entre eles, estão o fornecimento de água, de energia, de saneamento e de sistemas de transporte. Com isso, podemos avançar para a discussão acerca do desenvolvimento de projetos tecnológicos de infraestrutura na China.

Parte da estratégia chinesa com fins de desenvolver projetos de infraestrutura apoiados em Inteligência Artificial envolve o recrutamento de gigantes de tecnologia para comporem a equipe nacional de IA. Inicialmente, o time contava com três integrantes: o Baidu, o Alibaba e o Tencent. Cada empresa foi direcionada a realizar pesquisas e desenvolver produtos com base nas esferas em que possuíam experiência. O Alibaba, companhia que lidera o *e-commerce* chinês, foi escolhido para conduzir o desenvolvimento de tecnologias para as *smart cities*¹ (em português, cidades inteligentes). Mundialmente, uma de suas características mais marcantes é o Sistema de Transporte Inteligente, que integra o aparato físico à tecnologia e aos sistemas de informação para melhorar a experiência dos usuários.

Aqui cabe analisarmos com mais detalhes a infraestrutura associada ao transporte na China, um importante âmbito para o qual é destinada parte do capital nacional e do capital das *Big Techs* chinesas. Os investimentos em tecnologias aplicáveis à rede de transporte realizados na última década demonstram um direcionamento consonante com a estratégia nacional do país de levar prosperidade à sociedade. Isso porque sistemas de transporte mais eficientes com o apoio da tecnologia reduzem o tempo despendido no trânsito, na troca entre modais e na espera por transporte público, acarretando uma melhoria generalizada na vida da população.

Nesse contexto emergiu o conceito de “transporte inteligente”, conjuntura que caracteriza a rede de transporte de muitas cidades chinesas, mencionado anteriormente:

Transporte inteligente é a integração entre a tecnologia das informações

I *Smart city* (cidade inteligente) é um conceito que se refere a um ambiente urbano que utiliza dados e tecnologia a fim de aumentar a qualidade de vida dos cidadãos, otimizar o uso de recursos e proporcionar soluções urbanas (Cunha, 2016).

avançadas, comunicação de dados, transmissão de tecnologia, tecnologia do sensoriamento eletrônico, tecnologia do controle e tecnologia computacional em um sistema de gerenciamento de tráfego. A IA na indústria dos transportes na China vai além dos veículos individuais, e inclui interconectividade em toda a cidade e, possivelmente, a interconectividade em nível nacional. (Daxue, 2020, p. 95, tradução nossa).

Para ilustrar a tradução do investimento na tecnologia da IA em soluções inovadoras e inteligentes para o transporte, tem-se o *E.T. City Brain*, um sistema de gerenciamento de tráfego urbano alimentado pela nuvem e impulsionado pela Inteligência Artificial, desenvolvido para a cidade de Hangzhou pelo Alibaba (Alizia, 2018). A plataforma dá significado aos dados ao oferecer aos planejadores da cidade e autoridades de transporte uma maneira de analisá-los e transformá-los em informações operacionais em tempo real, utilizando-se de algoritmos avançados. A coleta do dado referente ao número de veículos na estrada, por exemplo, pôde resultar em uma redistribuição no trânsito que aumentou a velocidade média de viagem em cruzamentos congestionados.

Além disso, cabe citar também a solução de transporte inteligente “*TrafficGo*”, que está sendo desenvolvida pela empresa chinesa Huawei, *Big Tech* que foi incluída na equipe nacional de IA. A tecnologia em desenvolvimento se propõe a realizar um tratamento inteligente e automatizado de emergências em tempo real por meio da criação de um “cérebro humano” que administre e sincronize o tempo dos semáforos. Tal inovação tem como finalidade reduzir o atraso em vias principais — que implica no atraso em vias adjacentes — das metrópoles chinesas.

De acordo com a Huawei (2023), sem o uso do *TrafficGo*, as respostas às emergências são lentas, há uma falta de coordenação do sistema de sinais de trânsito, de otimização do tráfego e de integração espacial-temporal nas cidades. A criação da Huawei foi testada em áreas com um grande número de cruzamentos nas cidades de Shenzhen e Beijing e os resultados foram bastante positivos (Daxue, 2020). Conforme apresentou a empresa, o atraso médio na via principal da área-alvo do teste em Beijing reduziu em 25,2% e nas vias adjacentes, a redução foi entre 10 e 20%. A precisão dos sinais de

trânsito contribuem para um tráfego mais eficiente no qual a resolução de problemas se dá rapidamente.

Outro exemplo é o sistema de segurança inteligente nas estações de metrô implantado na cidade chinesa de Guangzhou. Essa solução se caracteriza pelo uso de um aplicativo de celular vinculado ao sistema de metrô em que o usuário se cadastra e realiza o seu escaneamento facial. Com isso, nas estações metroviárias, a verificação de segurança realizada no embarque se dá por meio do escaneamento facial do passageiro ou do *QR Code* gerado no App para o respectivo bilhete adquirido com o cadastro. Tal procedimento de verificação, marcado pela eficiência e precisão, leva menos de dois segundos para ser executado.

Já em Shanghai, as estações de trem implantaram um sistema de verificação de identidade ancorado também na tecnologia de reconhecimento facial, uma das aplicações da IA, que coordena a liberação da entrada nas estações. O escaneamento facial fornece informações referentes aos passageiros a um banco de dados que tem conexão com os cadastros dos usuários realizados no App em que são vendidos os bilhetes de trem. A liberação da entrada depende da equivalência entre o registro facial instantâneo e o registro facial cadastrado, atrelado ao banco de dados do sistema.

Nesse sentido, os projetos citados, caracterizados pelo protagonismo da Inteligência Artificial (IA), aumentam a eficiência de sistemas cujas respostas necessitam de rapidez, coordenação, dinamicidade, adaptabilidade e raciocínio. Ainda que os exemplos elencados apresentem uma escala local, é possível notar um caminho comum e um alinhamento entre eles, uma vez que a sua criação se pautou nos objetivos estratégicos nacionais da China apresentados tanto no NDAIDP quanto no 14º Plano Quinquenal.

Evidencia-se, portanto, a valorização e a pertinência do planejamento estatal no território chinês para a criação de projetos tecnológicos que contribuem com o desenvolvimento nacional, aproximando o país da prosperidade social. Pontua-se, por último, que o voluptuoso investimento em pesquisas para o desenvolvimento de tecnologias e para o avanço da teoria da IA têm feito a China assumir a liderança mundial na área.

Considerações finais

Buscou-se elucidar o processo de desenvolvimento chinês, com foco na tecnologia, a partir das reflexões propostas por Yuk Hui. Uma compreensão da cosmotécnica, conceito-chave, na medida em que o autor propõe um outro olhar para as diferentes realidades tecnológicas que existem no mundo atualmente e entende o solo oriental como fértil para o desenvolvimento dessa investigação.

A partir do exposto, entende-se que a China naturalmente está inserida em um sistema global onde impera o capitalismo e, por esse motivo, o país, ainda que seja regido pelo socialismo de mercado, está incluído numa lógica competitiva de liderança mundial. Diante disso, incorporou investimentos a fim de se tornar uma potência e, conseqüentemente, travar um embate — para além da ideologia — com outros países que também o fazem, especialmente em relação aos Estados Unidos.

No entanto, afirma-se que a preocupação da nação chinesa com o bem-estar comum e com o desenvolvimento de políticas que o promovam, dentro de uma estratégia nacional que engloba tanto as áreas urbanas quanto rurais, está relacionada ao histórico do pensamento tradicional chinês. Ainda que, em dado momento da história, a tradição tenha perdido protagonismo, percebe-se o enraizamento desse pensamento na cultura do país e, mais recentemente, no Partido Comunista Chinês e seu modelo de gestão e governança. Prova disso são os esforços de Xi Jinping para introduzir em seus discursos e políticas a influência do pensamento tradicional, principalmente do confucionismo.

O presente artigo é também uma investigação ainda em processo acerca dos acontecimentos da China contemporânea e da relação de sua ascensão com as tecnologias avançadas, como a IA. Yuk Hui é importante nesse processo, pois contribui com a expansão do pensamento e com o delineamento de relações virtuosas entre a filosofia e a técnica. Procurou-se, ainda, favorecer a construção de um outro olhar do Ocidente em relação ao Oriente, principalmente em relação à nação chinesa, entendendo que seu caminho até a contemporaneidade, apoiado numa cosmotécnica própria, é fonte de inspiração para outros países em desenvolvimento.

Referências

- ALIZIA. Alibaba Cloud making deeper push into smart cities. *Computer Weekly*, 2018. Disponível em: <<https://www.alizila.com/alibaba-cloud-launched-city-brain-2-o-hangzhou/>>. Acesso em: 13 ago. 2023.
- ASPI (Australian Strategic Policy Institute). *Critical Technology Tracker*, 2023. Disponível em: <<https://techtracker.aspi.org.au/>>. Acesso em: 8 ago. 2023.
- CHINA. Conselho Estadual. *Plano de Desenvolvimento de Inteligência Artificial de Nova Geração*, 2017. Disponível em: <https://www.gov.cn/gongbao/content/2017/content_5216427.htm>. Acesso em: 9 ago. 2023.
- CHINA. Conselho Estadual. *China's 2023 economic work priorities to facilitate high-quality development*, 2023. Disponível em: <http://english.www.gov.cn/news/topnews/202303/10/content_WS640a6f01c-6d0a757729e7e88.html#:~:text=Pursuing%20high%2Dquality%20development%20means,at%20Ren-min%20University%20of%20China>. Acesso em: 14 ago. 2023.
- CHINA. Conselho Estadual. *President Xi stresses innovation to bolster economy*, 2016. Disponível em: <http://english.www.gov.cn/news/top_news/2017/09/28/content_281475890616310.ht>. Acesso em: 11 ago. 2023.
- CHINA. Agência de Notícias Xinhua. *Décimo Quarto Plano Quinquenal para o Desenvolvimento Econômico e Social Nacional da República Popular da China e Esboço de Metas de Longo Prazo para 2035*, 2021. Disponível em: <https://www.gov.cn/xinwen/2021-03/13/content_5592681.htm>. Acesso em: 14 ago. 2023.
- CLAYTON, K. *Confucius technology and power consolidation under Xi Jinping*. *The China Story*, 2020. Disponível em: <<https://www.thechinastory.org/confucius-technology-and-power-consolidation-under-xi-jinping/>>. Acesso em: 13 ago. 2023.
- CUNHA, Maria Alexandra. *Smart Cities: transformação digital de cidades*. São Paulo: Programa Gestão Pública e Cidadania - Pgp, 2016. Disponível em: <<https://bibliotecadigital.fgv.br/dspace/handle/10438/18386>>. Acesso em: 12 ago. 2023.
- DAXUE CONSULTING. *The AI Ecosystem in China*, 2020. Disponível em: <<https://www.hhs.se/contentassets/bc962221471a415ba8aco1fbbf160277/chinas-ai-ecosystem-nov-2022.pdf>>. Acesso em: 8 ago. 2023.
- DOBBERSTEIN, L. *China leads the world in tech research, could win the future, says think tank*. *The Register*, 2023. Disponível em: <https://www.theregister.com/2023/03/03/china_leads_tech_aspi/>. Acesso em: 13 ago. 2023.
- HUAWEI. Huawei Cloud. *TrafficGo*, 2023. Disponível em: <<https://www.huaweicloud.com/product/trafficgo.html>>. Acesso em: 9 ago. 2023.
- HUI, Y. *The Question Concerning Technology in China: An Essay in Cosmotronics*. Londres: Urbanomic Media, 2016.

HUI, Y. *On the Persistence of the Non-modern*. Afterall, 2021. Disponível em: <<https://www.afterall.org/articles/on-the-persistence-of-the-non-modern/>>. Acesso em: 10 ago. 2023.

JABBOUR, E; GABRIELE, A. *China: o Socialismo do Século XXI*. São Paulo: Boitempo, 2021.

JABBOUR, E; DANTAS, A; BELMONTE, A. Ciclos econômicos, desenvolvimento e mercado: anotações sobre o socialismo e a 'etapa primária'. In: BERTOLINO, O; MONTEIRO, A. (Org.). *100 anos da Revolução Russa: legados e lições*. São Paulo: Anita Garibaldi, 2017.

JABBOUR, E.; DANTAS, A. Sobre a China e o “socialismo de mercado” como uma nova formação econômico-social. *Nova Economia*, [S.L.], v. 30, n. 3, p. 1029-1051, dez. 2020. *FapUNIFESP (SciELO)*. <http://dx.doi.org/10.1590/0103-6351/5544>. Disponível em: <<https://www.scielo.br/jj/neco/a/FrQC4MW-Z3W8hy5xLZkPQjmG/abstract/?lang=pt>>. Acesso em: 14 ago. 2023.

JABBOUR, E.; DANTAS, A.; VADELL, J. Da nova economia do projeto à globalização instituída pela China. *Estudos Internacionais: revista de relações internacionais da PUC Minas*, v. 9, n. 4, p. 90-105, 10 jan. 2022. Disponível em: <<http://periodicos.pucminas.br/index.php/estudosinternacionais/article/view/25823>> . Acesso em: 11 ago. 2023.

JABBOUR, E.; MOREIRA, W.; Do sistema nacional de inovação tecnológica à “Nova Economia Do Projeto” na China. *Brazilian Journal of Political Economy*, vol. 43, no 3, pp. 543-563, July-September/2023. Disponível em: <<https://doi.org/10.1590/0101-31572023-3455>> Acesso em: 11 ago. 2023.

MÁO JR., J. R. 60 ANOS DA REVOLUÇÃO CHINESA E 50 ANOS DA REVOLUÇÃO CUBANA: REVOLUÇÃO E LIBERTAÇÃO NACIONAL. *Revista Mouró*, [S. l.], v. 1, n. 2, 2010. Disponível em: <https://revista.mouro.com.br/index.php/Revista_Mouro/article/view/36>. Acesso em: 13 ago. 2023.

NEEDHAM, K. China leads tech race, highlighting need for AUKUS sharing, says think-tank. *Reuters*, 2023. Disponível em: <<https://www.reuters.com/world/china-leads-tech-race-highlighting-need-aukus-sharing-says-think-tank-2023-06-05/>>. Acesso em: 13 ago. 2023.

SERENI, E. De Marx a Lênin: a categoria de “formação econômico-social”. *Meridiano – Revista de Geografia*, n. 2, p. 248-346, 201.

SCIULO, M. M. 5 pilares centrais do pensamento de Confúcio. *Revista Galileu*, 2020. Disponível em: <<https://revistagalileu.globo.com/Sociedade/noticia/2020/01/5-pilares-centrais-do-pensamento-de-confucio.html>>. Acesso em: 13 ago. 2023.

Singularity Vs. Daoist Robots. *Noema Magazine*, 2020. Disponível em: <<https://www.noemamag.com/singularity-vs-daoist-robots/>>. Acesso em: 13 ago. 2023.

The Dexterous Butcher – Zhuangzi. *The Dewdrop*, 2020. Disponível em: <<https://thedewdrop.org/2020/05/18/the-dexterous-butcher-zhuangzi/>>. Acesso em: 10 ago. 2023.

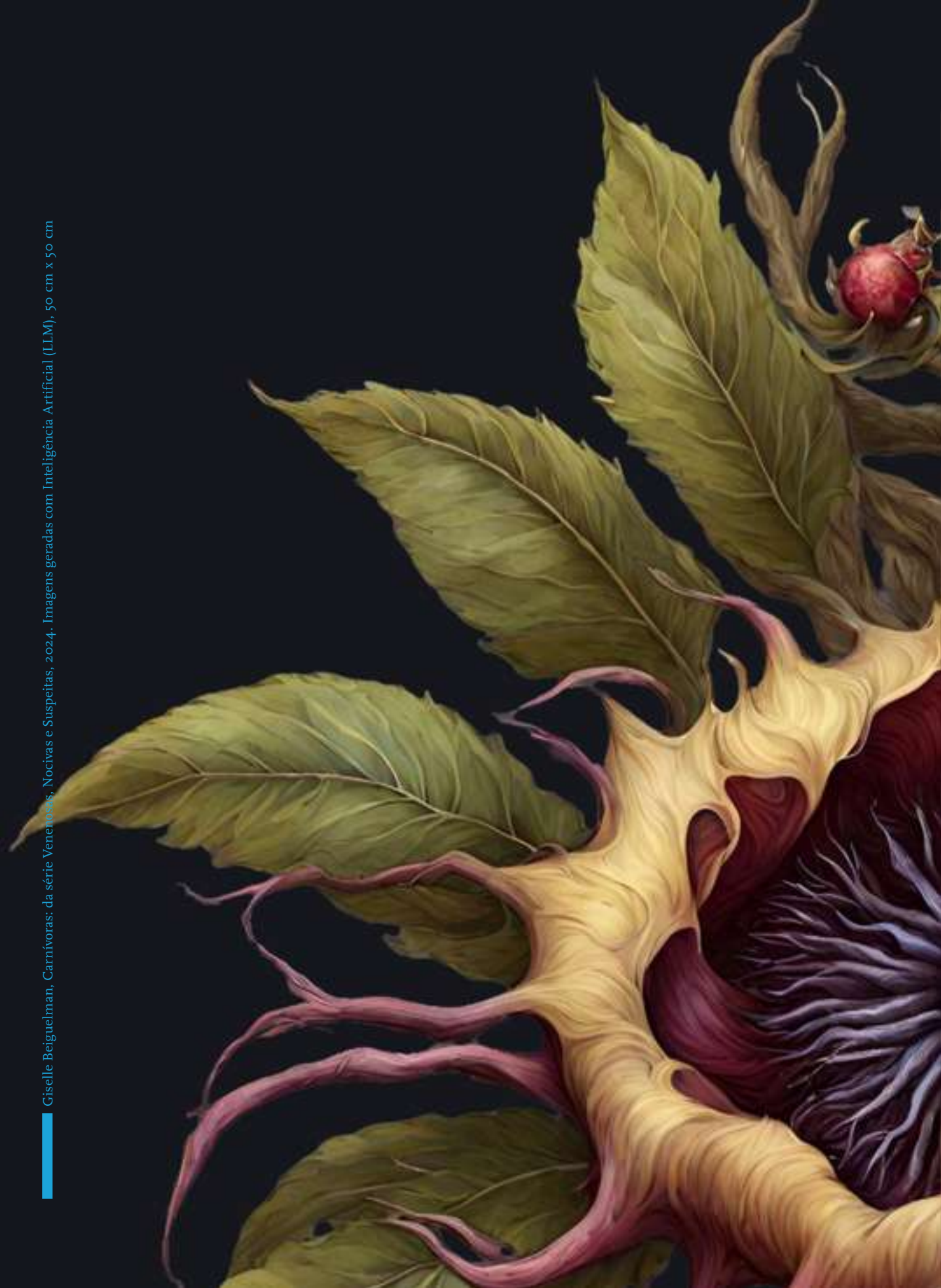
Yuk Hui e a pergunta pela cosmotécnica. Instituto Humanitas Unisinos, 2020. Disponível em: <<https://www.ihu.unisinos.br/categorias/602804-yuk-hui-e-a-pergunta-pela-cosmotecnica>>. Acesso em: 10 ago. 2023.



■ Giselle Beiguelman, frames do vídeo generativo da série Flora Rebellis, versão 2024.
Imagens geradas com Inteligência Artificial (Style Gans)

■ [Acesse o vídeo clicando na imagem](#)





INTELIGÊNCIA ARTIFICIAL E DIREITO:

o impacto transformador na organização de grandes acervos de textos jurídicos.

DOI: <https://doi.org/10.35699/2965-6931.2023.47689>

DANIELA VIANNA*; (JUSBRASIL) MARINA SOARES MARINHO**;
TIAGO FERREIRA***; ALTIGRAN SOARES DA SILVA****

RESUMO: Os recentes avanços na área de inteligência artificial e de processamento de linguagem natural têm impulsionado diversas mudanças no campo jurídico. No Brasil, o Direito segue um movimento constante de modernização, principalmente direcionado à transparência e ao acesso à informação. O grande volume de documentos jurídicos abre espaço para o desenvolvimento e uso de ferramentas inteligentes que busquem organizar e facilitar a gestão desse acervo. Neste trabalho, mostramos como o uso de modelos de linguagem em conjunto com técnicas de modelagem de tópicos são capazes de organizar e extrair conhecimento desses grandes acervos jurídicos, revelando temas que muitas vezes são implícitos e desconhecidos, o que traz benefícios para diversas aplicações, como a pesquisa por documentos similares e a recomendação de textos jurídicos.

PALAVRAS CHAVE: processamento de linguagem natural; modelos de linguagem; modelagem de tópicos; direito; organização de acervos jurídicos. abstract: Debates on technological advances have raised several issues over the past

Artificial Intelligence and law: the transformative impact on the organization of large collections of legal texts.

ABSTRACT: Recent advances in the area of artificial intelligence and natural language processing have driven several changes in the legal field. In Brazil, Law follows a constant movement of modernization, mainly aimed at transparency and access to information. The large volume of legal documents opens up space for the development and use of intelligent tools that seek to organize and facilitate the management of this collection. In this work, we show how the use of language models in conjunction with topic modeling techniques are capable

KEYWORDS: natural language processing; language models; topic modeling; right; organization of legal collections.

*JusBrasil/ ** Universidade Federal de Minas Gerais (Jusbrasil); ***Universidade de Brasília (Jusbrasil);
**** Universidade Federal do Amazonas

Introdução

Inteligência Artificial (IA) é a área de estudo dedicada a criar sistemas que sejam capazes de replicar e potencializar tarefas realizadas tipicamente por humanos. Um dos seus principais motores de desenvolvimento é o aprendizado de máquina (machine learning), uma sub-área da IA que permite que as máquinas aprendam padrões e tomem decisões a partir de dados. Embora as aplicações que utilizam inteligência artificial com base no aprendizado de máquina não sejam novas, é inegável a sua expansão nos últimos anos. Parte desse movimento pode ser atribuída ao aumento da quantidade de dados disponíveis, mas não apenas. Houve aumento expressivo na capacidade de processamento desses dados.

No campo da linguagem escrita, que nos interessa para tratar do Direito, os avanços mais notáveis ocorreram na área do processamento de linguagem natural (PLN - natural language processing, NLP). PLN é uma aplicação da IA que permite que a máquina compreenda a linguagem humana, permitindo a extração e a geração de textos. Em sua versão mais moderna, se vale de representações vetoriais (embeddings) para mapear as palavras a partir de um modelo de linguagem, como é o caso dos modelos Word2Vec (Mikolov et al., 2013), Doc2Vec (Le; Mikolov, 2014), Bert (Devlin et al., 2019b) e GPT (Brown et al., 2020). Estes, por sua vez, utilizam estatística e probabilidade para prever palavras de um texto a partir do seu contexto. Eles se baseiam em redes neurais artificiais, que visam a simular o processo de pensamento humano e a forma como processamos a linguagem. Esses modelos “adquirem conhecimento” por meio do treinamento em grandes corpora de texto, acumulando informações sobre como as palavras são utilizadas em diferentes tipos de documentos. Atualmente, os grandes modelos de linguagem são capazes de realizar diversas tarefas de processamento de linguagem natural como por exemplo a tradução automática, a sumarização e a geração de textos.

Justamente pela característica predominante da linguagem escrita, os últimos avanços integrando modelos de linguagem e processamento de linguagem natural são promissores para o meio jurídico. O enorme volume de documentos com os quais os acadêmicos e os profissionais da área lidam abre espaço para que ferramentas que

buscam organizar e facilitar a gestão desse acervo anunciem verdadeiras revoluções. Não seria a primeira vez. O mercado jurídico e até o Poder Judiciário seguem um movimento constante de modernização e investimento no acesso facilitado à informação. O desenvolvimento tecnológico permitiu, por exemplo, que o conteúdo dos autos processuais migrasse da Secretaria para as sedes dos escritórios de advocacia e até mesmo a residência dos jurisdicionados. A busca por jurisprudência, antes limitada ao acesso a dezenas de decisões, saiu das páginas de periódicos para a consulta rápida a uma base de milhões de documentos, filtrada a partir de palavras-chave. O processo de reconhecimento de entidades permitiu que as citações e intimações sejam mapeadas e informadas a partir de alertas por emails. Para mais, a tecnologia também permitiu que mesmo durante o isolamento provocado por uma pandemia global, por aproximadamente dois anos, a Justiça brasileira não apenas seguisse funcionando, como ampliasse a sua capacidade de julgamento.

Neste artigo vamos descrever um dos recentes casos de uso que desenvolvemos utilizando IA aplicada ao contexto do Direito: a utilização de modelos de linguagem e técnicas de modelagem de tópicos para a organização de grandes acervos de textos jurídicos.

Na primeira seção deste trabalho, Inteligência Artificial - Visão Geral e Estado da Arte do Processamento de Linguagem Natural, realizamos uma revisão abrangente das tecnologias e métodos mais recentes em IA, com uma ênfase especial naquelas que são relevantes para o campo do Direito e a modelagem de tópicos. Destacamos as principais pesquisas e avanços tecnológicos recentes nessa área e também identificamos as limitações atuais dessas tecnologias. Essa análise do estado da arte proporciona um contexto fundamental para se familiarizar com os projetos executados no campo jurídico, apresentados na seção seguinte, e para compreender o estudo de caso apresentado ao final, demonstrando assim a contribuição significativa e relevante que ele traz para a área do Direito.

Já na seção seguinte, A Utilização da IA no Direito, analisamos com mais detalhes os impactos recentes da IA para o Direito, a partir da discussão sobre as aplicações práticas da IA. Abordamos principalmente o mencionado processo de modernização da gestão da informação no campo jurídico e a estreita relação entre o desenvolvimento tecnológico e a viabilização do acesso à Justiça, um direito fundamental, constitucionalmente garantido.

Finalmente, na terceira seção, Estudo de Caso: Organização de Documentos Jurídicos, apresentamos um estudo de caso no qual investigamos o emprego de técnicas de tipificação visando à organização não supervisionada de um extenso volume de textos jurídicos. Inicialmente, descrevemos o desafio que essa abordagem visa superar, detalhando a implementação das técnicas utilizadas e os dados empregados. Em seguida, analisamos os resultados obtidos no estudo de caso, considerando suas inferências e comparando-os com a atualidade do estado da arte. A partir deste trabalho, conseguimos identificar conteúdo jurídico relacionado por temas mais gerais e mais específicos, o que permite recuperar documentos com maior precisão. Concluímos o estudo de caso ao explorar suas implicações potenciais, futuras aplicações e destacando a necessidade de pesquisas adicionais neste campo específico.

Inteligência artificial - visão geral e estado da arte do processamento de linguagem natural

Quando Alan Turing propôs o famoso teste que levou seu nome, o objetivo foi criar uma forma de atestar a capacidade de pensamento de uma máquina. O critério de aprovação era simples: se a máquina conseguisse emular um diálogo a ponto de ser indistinguível de um ser humano, essa máquina teria passado no teste (Russel, 2010). Tal capacidade conversacional é hoje um campo inteiro dentro da Inteligência Artificial, chamado de Processamento de Linguagem Natural (PLN) (Jojri, 2021).

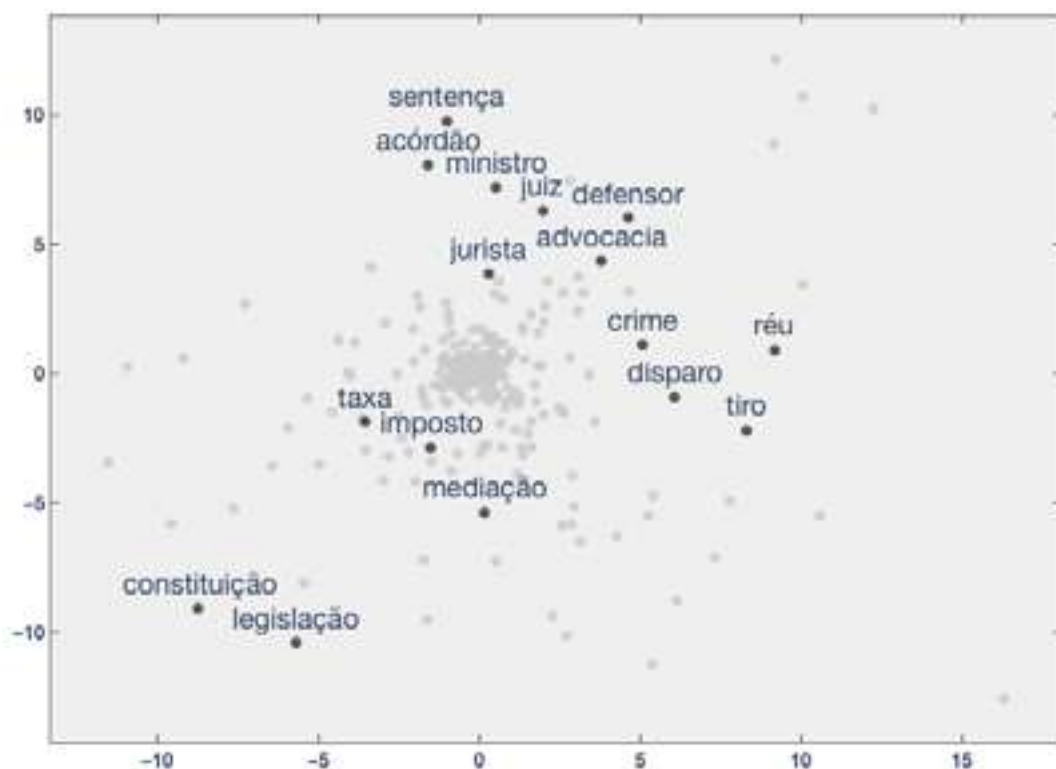
Esse campo surgiu da combinação de esforços entre a Linguística e a Ciência da Computação para permitir a criação de programas de computador que pudessem se comunicar de forma fluida com seres humanos (Jojri, 2021). Atualmente, desdobramentos diretos e indiretos do PLN são ubíquos, podendo ser encontrados em sistemas de busca, aplicativos de navegação, serviços de e-mail, editores de texto, e os inúmeros assistentes virtuais disponíveis em smartphones ou aparelhos eletrônicos em nossas casas. O PLN é composto por várias áreas, tais como Reconhecimento de Fala, Classificação de Texto, Tradução Automática, Autocorreção, Question Answering², Sumarização e Modelagem de Tópicos — assunto deste artigo (Dieng, 2020; Jojri, 2021). Apesar de serem empregadas com finalidades distintas, todas essas áreas têm como objetivo

projetar e construir sistemas que sejam capazes de entender a linguagem natural, falada ou escrita, ou de gerar conteúdo também em linguagem natural.

Em suas primeiras décadas, o PLN foi abordado de forma mais morfológica, ou seja, com regras rígidas e uma forte conexão com o campo da linguística. Isso requeria um profundo conhecimento da língua e definições inequívocas de regras sintáticas (Jojri, 2021). Mais tarde, com a evolução do poder computacional, essa abordagem foi gradualmente complementada com técnicas estatísticas, permitindo o desenvolvimento dos chamados “word embeddings”, que são representações vetoriais baseadas na probabilidade da ocorrência das palavras de um vocabulário a partir do contexto onde elas ocorrem (Liu, Q, 2020).

Em sua essência, embeddings podem ser visualizados como palavras distribuídas num plano cartesiano, onde a proximidade espacial entre palavras diferentes também significa uma proximidade semântica entre elas (Lund, 1996). Os embeddings utilizados atualmente podem possuir centenas e até milhares de dimensões, mas para uma visualização simplificada, podemos utilizar uma técnica de redução de dimensionalidade para representá-los em duas ou três dimensões (Figura 1).

Figura 1 - Representação gráfica de word embeddings em duas dimensões Fonte: elaborado pelos autores



Uma vez sendo possível representar textos de forma estritamente matemática, o PLN passou por um ponto de inflexão e avançou substancialmente o estado da arte na área (Hofmann, 2020). Com suas representações vetoriais, passamos a analisar palavras e textos aplicando a Álgebra Linear, e então simples analogias como “Rei - Homem + Mulher = Rainha” e “Acórdão - Ministros + Juiz = Sentença” passam a ser demonstráveis com um rigor matemático (Figura 2). E além de elementos em operações aritméticas, textos agora também podem ser usados como parâmetros em funções.

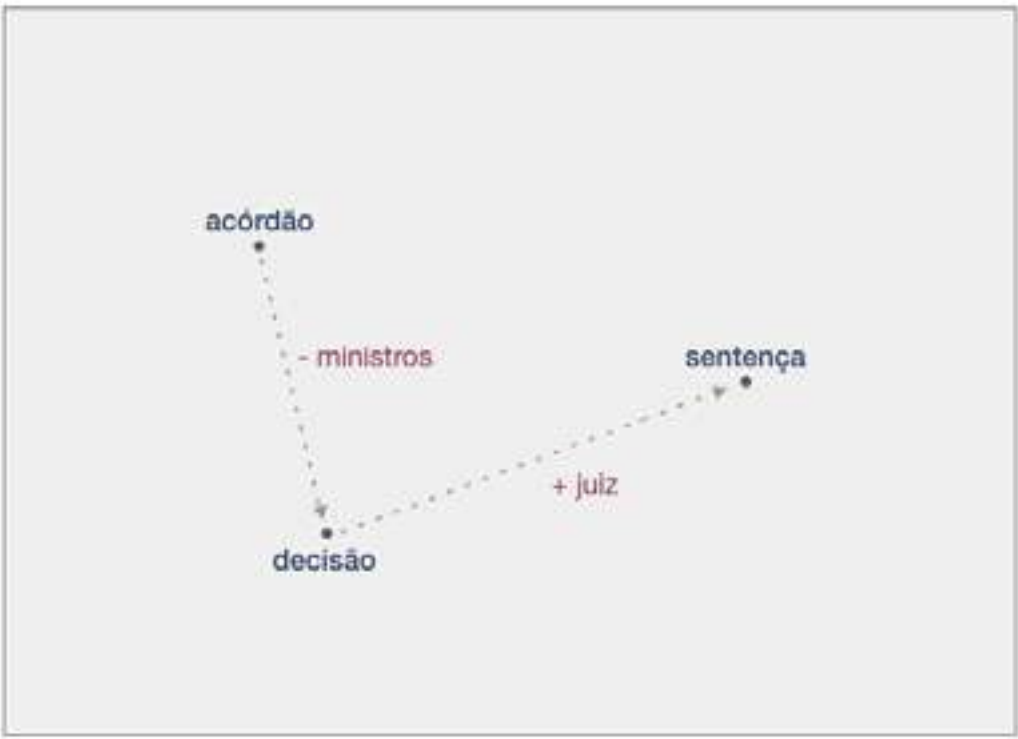


Figura 2 - Representação de operações aritméticas baseadas em word embeddings

Fonte: elaborado pelos autores

Até 2018, as técnicas de geração de word embeddings mais comuns, dentre elas Word2Vec e Doc2Vec, eram as chamadas “estáticas” e não tinham a capacidade de lidar com polissemia, homonímia e contextos diferentes (Wiedmann, 2019). Por exemplo, com essas técnicas, a palavra “pena” teria a mesma representação vetorial nos trechos “a pena foi fixada em 2 anos” e “o réu teve pena da vítima”. Foi então com o surgimento dos chamados word embeddings contextualizados, como o Bert (Delvin, 2019), que a relação entre as palavras e seu contexto passou a ser levada em consideração na geração de cada embedding, permitindo um aumento na performance de várias tarefas do

PLN (Liu, Q, 2020). Assim, a área passou por mais um ponto de inflexão (Ethayarajh, 2020), e essa nova arquitetura abriu caminho para aplicações mais robustas que, apesar de requerer mais recursos computacionais (ARORA, 2020), avançou mais uma vez o estado da arte (Wiedmann, 2019).

Especialmente nos últimos meses, o PLN tem ganhado destaque na mídia devido aos avanços nos Grandes Modelos de Linguagem e o lançamento do ChatGPT (Bartholomew, 2023), um chatbot tão sofisticado que não só passaria com sucesso no Teste de Turing (Biever, 2023), como já vem sendo utilizado profissionalmente em diversas áreas, incluindo a prática do Direito (Schwaecz, 2023). Levantando questões sobre privacidade, direitos autorais, geração de conteúdo malicioso e impactos no mercado de trabalho, a evolução da performance no PLN é tão fascinante quanto é alarmante e nos desafia a enfrentar as implicações tecnológicas, éticas e econômicas da adoção da IA em nossa sociedade.

No Direito, as aplicações de IA já permeiam boa parte das atividades relacionadas à prática da profissão. Desde tarefas de revisão de textos, como contratos e peças processuais, até pesquisa jurídica e análise preditiva, softwares e plataformas se utilizam das diversas técnicas de aprendizado de máquina e PLN para automatizar e acelerar o trabalho de advogados, magistrados e operadores do Direito. Na seção seguinte analisamos com mais detalhes algumas dessas aplicações.

3 Entre eles, destaca-se o LEIA, realizado pela empresa privada Softplan, o chatbot do TRE-DF, desenvolvido em parceria com o Whatsapp e o Vitor, do STF, construído em parceria com a Universidade de Brasília.

A utilização da ia no direito

De acordo com o Painel de Projetos com Inteligência Artificial no Poder Judiciário, o relatório mais detalhado, publicado pelo Conselho Nacional de Justiça (CNJ), estavam em curso no país, em 2020, 41 projetos envolvendo a IA aplicada ao Direito, distribuídos entre 32 Tribunais (Brasil, 2020). Destes, a maioria pertencia a Tribunais Estaduais (19) e foi desenvolvida inteiramente pela equipe técnica de suas instituições (24), mas não apenas: há relatos de projetos executados por inteiramente ou em parte por empresas privadas³ e em parceria com Universidades. A maioria dos projetos foi realizado em Python (36) e aplica métodos de redes neurais e redes neurais convolucionais (Convolutional Neural Networks, CNN) (16), além de árvores de decisão e floresta aleatória (12) (Brasil, 2020).

4 O chatbot Bel foi premiado na categoria Inovação Tecnológica no Prêmio de Inovação Judiciário Exponencial, em 2021 (Brasil, 2021a).

5 O nome Celina homenageia à natalense Celina Guimarães, primeira mulher a exercer o direito de voto no Brasil.

6 O LEIA também possui um módulo de precedentes, adotado pelo Tribunal de Justiça do Mato Grosso do Sul (TJ/MS). Este módulo auxilia Poder Judiciário na identificação de possíveis candidatos à vinculação à um dos temas de precedentes em tramitação nos Tribunais Superiores. Com base em matrizes de entendimento criadas pela empresa e validadas pelo Núcleo de Gerenciamento de Precedentes do TJ/MS (NUGEP/TJMS), a aplicação utilizou técnicas de processamento de linguagem natural e automação de consultas em base indexada baseado em regras definidas por seres humanos para realizar a “leitura e interpretação” da petição inicial do processo e verificar se está relacionada à um dos temas selecionados. (Brasil, 2019)

7 O Toth, criado pelo Tribunal de Justiça do Distrito Federal e dos Territórios (TJDFT) também é um classificador, criado para recomendar a classe e os assuntos do processo, durante a etapa do fluxo do PJe, identificada como petição inicial. O Tribunal

Entre as ferramentas, destacavam-se desde já os chatbots e assistentes virtuais. Por exemplo, o Bot Eleitoral (BEL) do Tribunal Regional Eleitoral do Espírito Santo (TRE/ES), um dos projetos descritos no Painel, é um assistente virtual que utiliza processamento de linguagem natural para responder a um menu de perguntas usuais feitas ao Disque Eleições, relativas ao número de título eleitoral, local de votação, situação eleitoral, justificativa etc.⁴ De forma semelhante funciona a Celina, um assistente virtual construído pelo TRE de Rondônia (TER/RN), disponível no site oficial do Tribunal ou no Telegram para responder questões sobre local de votação e dúvidas sobre os serviços prestados no Tribunal.⁵

Além das ferramentas criadas para interagir diretamente com a usuária, respondendo a perguntas, a mais numerosa das iniciativas desenvolvidas pelos Tribunais brasileiro para lidar com grandes volumes de documentos é o classificador de documentos de um processo. É o caso do Hércules, construído pelo Tribunal de Justiça do Estado do Alagoas (TJ/AL), a ferramenta propõe uma classificação automática de petições intermediárias na 15ª vara de execuções fiscais de Maceió e a posterior distribuição por filas de acordo com a classificação obtida. Além disso, há a validação do preenchimento de Certidões de Dívida Ativa, considerando os requisitos legais. No mesmo sentido é o LEIA Petições Intermediárias, criada por uma empresa do mercado privado, mas implementada pelo TJAL e pelo Tribunal de Justiça do Estado do Amazonas (TJ/AM) e que também promove a classificação automática de petições intermediárias, a partir da análise, identificação e sugestão do tipo e da categoria da petição.⁶ Ambos os projetos de classificação se propõem a facilitar a gestão da documentação jurídica e garantir celeridade na tramitação processual.

Também já utilizavam ferramentas dedicadas a classificar documentos o Tribunal de Justiça do Estado do Amapá (TJ/AM), a partir do Tucujuris Inteligência Artificial (TIA); o Tribunal de Justiça do Estado do Goiás (TJ/GO), com o Berna; o Tribunal de Justiça do Estado do Pernambuco (TJ/PE), que adota a Elis; o Tribunal de Justiça do Estado do Paraná (TJ/PR), a partir do Robo Larry; o Tribunal de Justiça do Estado de Santa Catarina (TJ/SC), com o Classificador de Petições de Execução Fiscal; o Tribunal de Justiça do Estado do Tocantins, com o MinerJus; o Tribunal Federal da 1ª Região (TRF1), que implementou o Análise Legal Inteligente (ALEI); o Tribunal Federal da 4ª Região (TRF4), que agrupa apelações por similaridade de sentenças; o Tribunal Federal

da 5ª Região (TRF5), o qual agrupa petições iniciais para realização de perícia médica; o Tribunal Superior do Trabalho (TST), que criou o Bem-te-vi para triar documentos processuais; o Tribunal do Trabalho da 7ª Região (TRT7) e da 20ª Região (TRT20), que agrupa documentos para facilitar a distribuição de trabalho entre funcionários públicos com o GEMINI.⁷

Entre as principais aplicações utilizadas em Tribunais Superiores, a IA Victor, recentemente remodelada como VictorIA, é um sistema que apoia a atividade de análise de admissibilidade recursal por meio da sugestão de um ou mais temas de repercussão geral, posteriormente sujeita à validação pelos servidores e pelos Ministros (Salomão, 2023, p. 27-31). A análise de admissibilidade é o procedimento de verificar se todos os requisitos para propor um recurso no STF foram atendidos; quando um recurso é muito similar a outro anteriormente proposto, aquele deve aguardar o julgamento deste. Com os avanços descritos em NLP, a VictorIA passou a agrupar processos por similaridade de temas, de forma que consegue identificar novas controvérsias (Brasil, 2023).

De acordo com o Segundo e o Terceiro Relatórios “Tecnologias Aplicadas à Gestão de Conflitos no Poder Judiciário com ênfase no uso da inteligência artificial”, coordenado pelo Ministro Luis Felipe Salomão, do Superior Tribunal de Justiça (STJ), em 2021 o mapeamento de iniciativas de IA aplicadas ao Direito no Poder Judiciário identificou 64 ferramentas do tipo, 23 a mais que no ano anterior. Já em 2022, foram identificados 85 novos projetos de IA.⁸ A conclusão dos Relatórios é de que a maioria das ferramentas tem a sua construção motivada pelo aumento de produtividade, a busca pela inovação, a melhoria da qualidade dos serviços judiciais e a redução de custos (Salomão, 2023). Objetivos relacionados com o princípio constitucional da eficiência aplicado ao Poder Público e com a garantia do acesso à justiça e do devido processo legal, expressamente destacados no art. 5º da Constituição da República Federativa do Brasil de 1988 (CFRB/1988).

A carta constitucional assegura, como direito e garantia fundamental dos brasileiros e estrangeiros residentes no país, a vida, a liberdade, a igualdade, a segurança e a propriedade (art. 5º). Faz isso especificando direitos e garantias individuais e coletivas, entre elas determinando que ninguém “será privado da liberdade ou de seus bens sem o devido processo legal” (art. 5º, inc. LIV); garantindo o “direito de petição

é pioneiro em aplicações utilizando IA. A época do lançamento do Painel, estavam em produção 4 projetos: além do Toth, o Amon (para implementar um sistema de reconhecimento facial a partir de fotografias), o Artiu (dedicado ao encaminhamento correto de mandados para distribuição e cumprimento nos endereços dos jurisdicionados), o Hórus (para inserir automaticamente no Projeto Judicial Eletrônico (PJe) os processos digitalizados e Natureza Conciliação (para atendimento aos Centros Judiciários de Solução de Conflitos e Cidadania – CEJUSCs a partir da classificação de procedimentos que gerassem uma ou mais sessões de conciliação) (Brasil, 2021b).

8 12 projetos identificados em 2021 foram cancelados ou suspensos [REF].

aos Poderes Públicos em defesa de direitos ou contra ilegalidade ou abuso de poder” e protegendo a coisa julgada (art. 5^ºffl, inc. XXIV) e assegurando que “a lei não excluirá da apreciação do Poder Judiciário lesão ou ameaça a direito” (art. 5^ºffl, inc. XXV).

Os mencionados incisos do art. 5^ºffl viabilizam a vida, a liberdade, a igualdade, a segurança e a propriedade porque promovem previsibilidade, confiabilidade e razoável rapidez na solução de conflitos. Trata-se de requisitos basilares do Estado Democrático de Direito (Waldron, 2008), os quais estruturam as relações sociais por meio do ordenamento jurídico. O julgamento parcial, as condições desiguais de acesso à Justiça, a demora na apreciação de pedidos litigiosos, muitas vezes implicando em privação de liberdade e danos materiais, não se conformam, de nenhuma maneira, ao desenho constitucional do Estado brasileiro. Democratizar o acesso à informação é garantir igualdade, dignidade e cidadania.

O investimento patente em classificadores e agrupadores de documentos por similaridade reflete a demanda de o Sistema Jurídico brasileiro prover condições de materialidade ao art. 5^ºffl da CRFB/1988. A partir da formação de grupos de conteúdo e documentos é mais fácil selecionar o mais adequado à necessidade informacional do jurisdicionado. Além disso, facilita a descoberta de novas oportunidades de atuação, o que promove equalização das condições de trabalho em um mercado tão desigual quanto o jurídico. Na seção seguinte, investigamos o emprego de técnicas de modelagem de tópicos e modelos de linguagem visando a organização e sumarização de textos jurídicos.

Estudo de caso: organização de documentos jurídicos

Investir em ferramentas inteligentes para apoiar especialistas jurídicos e acelerar decisões legais é crucial para muitos países devido à complexidade do sistema judiciário em todo o mundo. Nesse estudo de caso, focamos na tarefa desafiadora de organizar grandes e complexas coleções de documentos legais, sendo este um passo importante para a construção de soluções inteligentes eficientes para diversos problemas no domínio jurídico. A organização de documentos jurídicos serve de base para a realização

de diversas outras tarefas, como a busca/pesquisa de casos jurídicos similares (Shao et al., 2021; Sansone; Sperlí, 2022; Novaes et al., 2023), recomendação de artigos e livros jurídicos (Yang et al., 2021; Huang et al., 2021) e, classificação de texto jurídico (Chen et al., 2022; Chalkidis et al., 2019; Sulis et al., 2022).

Texto jurídicos são na maioria das vezes longos, escritos num jargão muito próprio e quase sempre de difícil compreensão para o público em geral. Além disso, um único texto pode tratar de uma variedade de temas diferentes. A tarefa de organizar textos jurídicos envolve o desenvolvimento de técnicas computacionais capazes de extrair conhecimento desses documentos fracamente estruturados e complexos, revelando temas, que muitas vezes são implícitos e desconhecidos, principalmente quando consideramos a natureza dinâmica do universo jurídico.

Nos últimos anos, com os avanços na área de PLN, mais especificamente em modelos de tópicos neurais (Dieng Ruiz; Blei, 2020; Bianchi; Terragni; Hovy, 2021; Srivastava; Sutton, 2017) e representações contextuais para texto (Devlin et al., 2019), as técnicas de descoberta e modelagem de tópicos foram aprimoradas e hoje podem ser aplicadas de forma eficaz a toda e qualquer coleção de documentos, incluindo documentos jurídicos com toda peculiaridade que esses apresentam em sua linguagem.

Nessa seção, mostramos a viabilidade do uso de técnicas modernas de descoberta de tópicos em conjunto com representações contextuais de texto para permitir a organização de uma coleção de documentos jurídicos composta por aproximadamente 14 mil acórdãos do STJ (Supremo Tribunal de Justiça) relacionados ao assunto “Contribuições Previdenciárias”. Acórdãos são decisões proferidas por um colegiado em tribunais de segunda instância, no caso desse estudo, o STJ. Na subseção seguinte, detalhamos esse estudo de caso, seguido de uma avaliação quantitativa e qualitativa da solução desenvolvida. Uma versão detalhada deste estudo foi apresentada em (Vianna et al., 2023).

Descoberta de tópicos para o domínio jurídico

No contexto das técnicas de modelagem e descoberta de tópicos, um tópico consiste em um agrupamento de palavras-chave que juntas possuem algum significado expressivo, permitindo a interpretação e organização de coleções de dados. Um algoritmo de descoberta de tópicos precisa analisar todos os documentos da coleção em busca de padrões de palavras e expressões que sejam semanticamente semelhantes e que melhor representem a temática de subconjuntos de documentos. Todo o processo é não supervisionado; ou seja, não existe nenhum conhecimento prévio dos tópicos que compõem a coleção de dados.

O estudo começa com a criação da coleção de dados a ser organizada. Para isso, expressões regulares contendo as palavras-chave ‘contribuições previdenciárias’ e ‘contribuição previdenciária’ são utilizadas para extrair, a partir de uma base de dados maior (Vianna et al., 2023), acórdãos do STJ que mencionam alguma variação desses termos. Em seguida, diversas técnicas de pré-processamento são aplicadas ao texto original com o intuito de limpar e normalizar a coleção. Entre as técnicas empregadas está a conversão de todo o texto para minúscula e a remoção de URLs, stopwords e pontuação. Essas duas etapas, extração e pré-processamento, resultaram em uma coleção contendo aproximadamente 14 mil acórdãos do STJ sobre o assunto “Contribuições Previdenciárias”. É importante ressaltar que essa coleção contém apenas acórdão publicados até 2022.

A partir da criação e pré-processamento da coleção de dados, um método de descoberta de tópicos, denominada BERTopic (Grootendorst, 2022) é aplicada a todos os documentos da coleção resultando na descoberta de aproximadamente 300 tópicos dentro do assunto principal ‘Contribuições Previdenciárias’.

BERTopic é um método de modelagem de tópicos baseada na descoberta de clusters (agrupamentos) densos de palavras a partir de representações vetoriais de documentos, denominados embeddings. Nesta técnica, documentos são representados no espaço vetorial de maneira que a similaridade semântica entre eles pode ser calculada. BERTopic parte do princípio de que documentos que abordam os mesmos tópicos são semanticamente similares. Uma vez que os documentos são projetados no espaço vetorial, uma técnica tradicional de aprendizado de máquina, denominada de redução de

dimensionalidade, é aplicada aos vetores que representam os documentos, com intuito de reduzir a complexidade de tais vetores e permitir a aplicação de outras técnicas de aprendizado de máquina não supervisionado, como é o caso dos algoritmos de agrupamento. Depois de reduzidos os vetores, um algoritmo de agrupamento, chamado Hdbscan (Rahman et al., 2016), é aplicado as representações vetoriais dos documentos. O objetivo é agrupar documentos relacionados em um mesmo cluster ou grupo. As palavras-chave que representam um tópico são extraídas a partir desses documentos usando uma técnica chamada de c-TF-IDF (Grootendorst, 2022), responsável por calcular a relevância de um termo a partir dos documentos de um tópico. É relevante mencionar, que o algoritmo HDBSCAN é uma técnica de agrupamento conhecida como soft-clustering, ou seja, um documento pode ‘pertencer’ a diferentes clusters (tópicos) com probabilidades diferentes. Neste estudo, vamos focar no cluster (tópico) dominante de cada documento. Cluster dominante é o cluster com maior probabilidade com relação a um determinado documento.

Como mencionado no parágrafo anterior, antes de serem agrupados, os documentos da coleção precisam ser convertidos em uma representação vetorial. Dada que a coleção adotada é formada por documentos longos, contendo em média 3041 palavras, decidimos por adotar o modelo Doc2Vec (Le; Mikolov 2014). Diferentemente dos modelos atuais baseados em Transformers (Vaswani et al., 2011) tais como Bert (Devlin et al., 2019b) e GPT (Brown et al., 2020), que sofrem com limitações com relação ao tamanho do dado de entrada, Doc2Vec é capaz de gerar uma representação vetorial para todo o documento sem a necessidade de particionar esses documentos em blocos menores de texto.

Experimentos

Nesta seção apresentamos os resultados obtidos, descrevemos a metodologia de avaliação aplicada e uma breve discussão sobre os impactos de tais resultados para o universo jurídico brasileiro.

Metodologia de avaliação. Os resultados obtidos durante este estudo foram avaliados por um time de especialistas do direito. Os especialistas julgaram a qualidade de um tópico descoberto pelo método de acordo com: (1) a qualidade das palavras-chave

9 As pessoas especialistas que ajudaram na definição dos tópicos de pesquisa e na avaliação dos documentos agrupados integravam, à época, a equipe de Legal Consultants do Jusbrasil (Isabela Franco e Letícia Leite) e de Legal Engineering, mais especificamente do time de pesquisa jurídica (Luiz Azevedo; Maria Flávia Freitas Ribeiro; Catarina Wodzik; Ana Laura Monteiro e Victória Hollanda).

geradas para representarem o tópico e; (2) a qualidade do relacionamento entre os documentos agrupados em um mesmo tópico e a relação entre os documentos e o conjunto de palavras-chave gerado. Com relação a qualidade das palavras-chave, os especialistas atribuíram a cada tópico analisado uma nota variando de 1 a 5, onde 5 representa um tópico de excelente qualidade. Ou seja, o conjunto de palavras-chave é composto por termos que além de trazerem uma ideia clara do tópico em questão, permitem também uma fácil distinção entre os demais tópicos gerados. Por exemplo, os termos ‘rural, comercialização, empregador rural, comercialização produção, rural pessoa, inconstitucionalidade, pessoa física, repristinação, declaração inconstitucionalidade, funrural’ obtiveram a nota máxima (nota 5), uma vez que permitiram que os especialistas prontamente identificassem o assunto principal desse tópico ‘Contribuição exigida do empregador rural pessoa física’. Já tópicos com notas mais baixas, são aqueles cujas palavras-chave não trazem clareza com relação aos assuntos presentes nos documentos desse tópico. Nesse caso, as palavras-chave são muito gerais e sem significado claro. Por exemplo, os termos ‘cofins, ipi, receitas, pis, financeiras, monetária, creditamento, correção monetária, capital próprio, correção’ foram considerados como sendo de baixa qualidade (nota 1), uma vez que os especialistas do direito não foram capazes de identificar o assunto principal discutido nesse tópico. Para os especialistas, os termos foram considerados bem gerais, englobando uma variedade de assuntos.

Para avaliar a qualidade dos documentos agrupados em um tópico, os especialistas consideraram o mínimo de 10 documentos e o máximo de 20 documentos por tópico.⁹ Para cada um dos documentos avaliados, os especialistas avaliaram se os documentos são relacionados entre si e se abordam a temática sugerida pelas palavras-chave que representam o tópico. Sendo assim, documentos foram marcados como relevantes ou não para os tópicos a que pertencem. Para garantir que os tópicos avaliados sejam representativos de toda coleção, uma amostra dos tópicos foi selecionada considerando um intervalo de confiança de 95% e uma margem de erro de 10%. No total, os especialistas avaliaram 163 tópicos dos 303 tópicos encontrados pelo algoritmo de modelagem de tópicos.

RESULTADOS

O primeiro passo para validar os resultados obtidos, foi a análise dos especialistas das palavras-chave encontradas pelo método de modelagem de tópicos para a amostra

contendo 163 tópicos. Seguindo a metodologia proposta, os especialistas deram notas entre 1 e 5 para cada um dos tópicos avaliados, onde 5 representa tópicos de excelente qualidade. O Quadro 1 apresenta o número total e percentagem de tópicos para cada uma das cinco notas possíveis. Podemos notar que 58% dos tópicos foram considerados como sendo de alta qualidade, ou seja, as palavras-chave representam de forma clara a temática do tópico e ainda ajudam a diferenciar o tópico analisado dos demais tópicos encontrados pelo modelo. Apenas 3% dos tópicos receberam notas 1 e 2, indicando que são raros os casos onde os tópicos são irrelevantes, pelo menos no que se refere a qualidade das palavras-chave.

Quadro 1 – Avaliação das palavras-chave da amostra de 163 tópicos dos 303 tópicos encontrados pelo algoritmo BERTopic para a coleção de dados sobre ‘Contribuições Previdenciárias’ Fonte: autoria própria.

Nota 1	Nota 2	Nota 3	Nota 4	Nota 5
2 (1%)	4 (2%)	16 (10%)	46 (28%)	95 (58%)

Ainda seguindo a metodologia de avaliação proposta, para a amostra de 163 tópicos foram avaliados entre 10 e 20 documentos de cada tópico, resultando em um total de 1882 documentos avaliados. Dentre eles, 1797 (95%) documentos foram considerados relevantes pelos especialistas. Os resultados indicam que além da clareza expressa pelas palavras-chave na descrição dos tópicos, a qualidade dos agrupamentos de documentos se mostra relevante, o que pode ter aplicação direta em diversos problemas da área, como a recuperação de documentos jurídicos semelhantes, um trabalho que quando feito de forma manual pode ser lento e custoso.

O Quadro 2 ilustra alguns dos tópicos avaliados pelos especialistas. Na primeira coluna, apresentamos as palavras-chave que representam o tópico e, na segunda coluna, apresentamos comentários deixados pelos especialistas sobre cada um desses tópicos.

Concluimos que os agrupamentos foram organizados conforme os documentos tratam dos mesmos temas de cauda longa, ou seja, assuntos mais específicos dentre os existentes em um grupo temático mais amplo. Por exemplo, todos os temas do Quadro

2 pertencem ao ramo do Direito Tributário, mais especificamente, ao tema mais específico “contribuições previdenciárias”, que são os tributos pagos ao INSS. Dentro deste tema de mais alto nível, conseguimos separar agrupamentos identificados por tipos de contribuições previdenciárias, como demonstrado na linha 1. Também identificamos agrupamentos ainda mais específicos, como no caso da linha 4, em que os documentos tratavam nomeadamente da Lei estadual de Minas Gerais que instituiu contribuições previdenciárias para custeio de serviços de assistência médica, ou das linhas 6 e 7, que discutem a incidência dessas contribuições sobre verbas remuneratórias de natureza salarial polêmica, quais sejam, o auxílio alimentação e o décimo terceiro. Neste caso específico [das linhas 6 e 7], é muito importante ressaltar a especificidade dos agrupamentos, já que os documentos discutem teses advocatícias muito similares, mas foi capaz de segregá-los de acordo com a parcela remuneratória discutida.

Ou seja, a modelagem de tópicos foi capaz de, diante de um conjunto pré-selecionado de documentos, sem qualquer direcionamento, identificar assuntos específicos que eram discutidos em documentos jurídicos.

Quadro 2 – Alguns exemplos de tópicos e palavras-chave encontradas pelo algoritmo de modelagem de tópicos. Fonte: autoria própria.

Tópico: Palavras-chave	Comentário
incra, destinada incra, destinada, extinta, funrural, agrária, domínio econômico, reforma agrária, domínio, social, seguridade, econômico, empresas, contribuição, urbana, seguridade social, reforma, empresa urbana, sesc, intervenção	Diferentes assuntos relacionados à contribuição para o Incra.
renúncia, desaposentação, aposentadoria, devolução valores, renúncia aposentadoria, devolução, jubilamento, valores recebidos, benefício, recebidos, concessão novo, segurado, posterior jubilamento, concessão, novo benefício, valores, previdenciário, posterior, novo posterior, novo	Devolução dos valores recebidos em razão de desaposentação
rural, prova material, prova, início prova, testemunhal, início, material, prova testemunhal, atividade, idade, atividade rural, rurícola, autora, aposentadoria, exercício atividade, aposentadoria idade, período, carência, familiar, autor	Tempo de serviço rural/contribuições não recolhidas
saúde, minas gerais, minas, estado minas, gerais, mg, serviços saúde, custeio, custeio serviços, ip-semg, serviços, agrg mg, estado, indébito, mg min, repetição, assistência saúde, contribuição custeio, servidores estado, indevida	Lei estadual que instituiu CP para custeio de serviços de assistência médica
rural, prova material, prova, início prova, testemunhal, início, material, prova testemunhal, atividade, idade, atividade rural, rurícola, autora, aposentadoria, exercício atividade, aposentadoria idade, período, carência, familiar, autor	Tempo de serviço rural/contribuições não recolhidas
alimentação, in natura, natura, auxílio alimentação, pat, auxílio, programa alimentação, alimentação trabalhador, trabalhador pat, in, pago, alimentação pago, programa, trabalhador, fornecida, empregador, salário in, transporte, pecúnia, alimentação	Incidência de CP sobre auxílio alimentação
separado, natalina, gratificação natalina, gratificação, salário, cálculo separado, separado gratificação, tributação separado, incidente gratificação, mês dezembro, aplicação separado, terceiro salário, décimo terceiro, décimo, mediante aplicação, cálculo, dezembro, terceiro, separado tabela, tabela	CP sobre 13º salário (gratificação natalina)

Conclusões

Neste artigo, apresentamos um caso de uso de inteligência artificial e processamento de linguagem natural aplicado ao contexto do Direito no Brasil. Precisamente, mostramos que a utilização de modelos de linguagem combinados com a modelagem não supervisionada de tópicos pode ser aplicada de forma eficiente e precisa na organização de grandes acervos de texto jurídico. Experimentos foram realizados com uma coleção de dados composta por aproximadamente 14 mil acórdãos do STJ sobre o assunto “Contribuições Previdenciárias”. Os resultados foram cuidadosamente avaliados por um grupo de especialistas (Vianna et al., 2023) que concluiu que 86% dos 163 tópicos avaliados eram de boa qualidade (notas 4 e 5), o que significa que, no geral, as palavras-chave dos tópicos representam de forma clara e satisfatória o conteúdo de cada um desses tópicos. Além disso, os especialistas também avaliaram a qualidade e coerência entre os documentos de um mesmo tópico. Os resultados obtidos foram expressivos, com 95% dos documentos avaliados considerados relevantes com relação a seus tópicos.

Com os avanços na área de processamento de linguagem natural, mais especificamente, com os grandes modelos de linguagem, é possível vislumbrar possíveis melhorias principalmente em relação à descrição dos tópicos gerados. Por exemplo, modelos gerativos podem ser usados para enriquecer o conjunto de palavras-chave que descrevem um tópico, trazendo mais contexto sobre o conteúdo dos tópicos e automaticamente facilitando o entendimento do mesmo por parte não apenas dos especialistas do direito, mas também do cidadão comum. Uma outra possibilidade é o uso de modelos de linguagem para a geração de resumos que descrevam os tópicos em uma linguagem mais próxima da linguagem humana.

Além do caso de uso apresentado, esse estudo traz uma revisão abrangente das tecnologias e métodos de inteligência artificial e processamento de linguagem natural, com foco em aplicações no contexto do Direito do Brasil.

Referências

- ARORA, S. et al. Contextual embeddings: When are they worth it? *arXiv preprint arXiv:2005.09117* (2020).
- BARTHOLOMEW, J. et al. How the media is covering ChatGPT. Disponível em: <https://www.cjr.org/tow_center/media-coverage-chatgpt.php> Acesso em 12/08/2023
- BIANCHI, Federico; TERRAGNI, Silvia; HOVY, Dirk. Pre-training is a hot topic: Contextualized document embeddings improve topic coherence. In: *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*. [S.l.]: Association for Computational Linguistics, 2021. p. 759–766.
- BIEVER, C. ChatGPT broke the Turing test — the race is on for new ways to assess AI. Disponível em: <<https://www.nature.com/articles/d41586-023-02361-7>> Acesso em 12/08/2023
- BROWN, T. B. et al. Language models are few-shot learners. In: *Proceedings of the 34th International Conference on Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates Inc., 2020. (NIPS'20). ISBN 9781713829546.
- CHALKIDIS, I. et al. Extreme multi-label legal text classification: A case study in EU legislation. In: *Proceedings of the Natural Legal Language Processing Workshop 2019*. Minneapolis, Minnesota: Association for Computational Linguistics, 2019. p. 78–87. Disponível em: <<https://aclanthology.org/W19-2209>>.
- CHEN, H. et al. A comparative study of automated legal text classification using random forests and deep learning. *Information Processing Management*, v. 59, n. 2, p. 102798, 2022. ISSN 0306-4573. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0306457321002764>>.
- BRASIL. Conselho Nacional de Justiça. Painel de projetos com inteligência artificial no Poder Judiciário, 2020. Disponível em: <<https://paineisanalytics.cnj.jus.br/single/?appid=29d-710f7-8d8f-47be-8af8-a9152545b771&sheet=b8267e5a-1f1f-41a7-90ff-d7a2f4ed34ea&lang=pt-BR&opt=ctx-menu,cursel>>. Acesso em: 14/08/2023.
- BRASIL. Tribunal Eleitoral Regional do Espírito Santo. Bel, a assistente virtual do TRE-ES, vence a categoria “Inovação Tecnológica” do Prêmio de Inovação Judiciário Exponencial. Comunicação, 06 out. 2021. 2021a. Disponível em: <<https://www.tre-es.jus.br/comunicacao/noticias/2021/Outubro/bel-a-assistente-virtual-do-tre-es-vence-a-categoria-inovacao-tecnologica-do-premio-de-inovacao-judiciario-exponencial>>. Acesso em: 14 out. 2023.
- BRASIL, Tribunal de Justiça do Amazonas. TJAM automatiza classificação de petições intermediárias no Portal e-SAJ. Imprensa, 19 dez. 2019. Disponível em: <<https://www.tjam.jus.br/index.php/menu/sala-de-imprensa/2387-tjam-automatiza-classificacao-de-peticoes-intermediarias-no-portal-e-saj>>. Acesso em: 14 ago. 2023.
- BRASIL. Tribunal de Justiça do Distrito Federal e Territórios. TJDF lidera número de projetos de Inteligência Artificial no Poder Judiciário. Janeiro de 2021b. Disponível em: <<https://www.tjdft.jus.br/institucional/imprensa/noticias/2021/janeiro/tjdft-e-o-tribunal-com-mais-projetos-de-inteligencia-artificial>>. Acesso em: 14 ago. 2023.
- BRASIL. Supremo Tribunal Federal. Ministra Rosa Weber lança robô Vitória para agrupamento e

classificação de processos, 17 mai. 2023. Disponível em: <https://portal.stf.jus.br/noticias/verNoticiaDetalhe.asp?idConteudo=507426&ori=1>. Acesso em: 14 ago. 2023.

DEVLIN, J. et al. BERT: Pre-training of deep bidirectional transformers for language understanding. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. [S.l.: s.n.], 2019. p. 4171–4186.

DEVLIN, J. et al. BERT: Pre-training of deep bidirectional transformers for language understanding. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, Minnesota: Association for Computational Linguistics, 2019. p. 4171–4186. Disponível em: <<https://aclanthology.org/N19-1423>>.

DIENG, Adj. B.; RUIZ, Francisco J. R.; BLEI, David M. Topic modeling in embedding spaces. *Transactions of the Association for Computational Linguistics*, MIT Press, Cambridge, MA, v. 8, p. 439–453, 2020.

ETHAYARAJH, K. BERT, ELMo, & GPT-2: How Contextual are Contextualized Word Representations? Disponível em: <<http://ai.stanford.edu/blog/contextual/>> Acesso em 12/08/2023

GROOTENDORST, Maarten. Bertopic: Neural topic modeling with a class-based tf-idf procedure. *CoRR*, abs/2203.05794, 2022. Citado na página 1.

HOFMANN, V. et al. Dynamic contextualized word embeddings. *arXiv preprint arXiv:2010.12684* (2020).

HUANG, Z. et al. Context-aware legal citation recommendation using deep learning. In: *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Law*. New York, NY, USA: Association for Computing Machinery, 2021. (ICAIL '21), p. 79–88. ISBN 9781450385268. Disponível em: <<https://doi.org/10.1145/3462757.3466066>>.

JOJRI, P. et al. Natural language processing: History, evolution, application, and future work. In: *Proceedings of 3rd International Conference on Computing Informatics and Networks: ICCIN 2020*, pp. 365–375. Springer Singapore, 2021.

LE, Quoc; MIKOLOV Tomas. Distributed representations of sentences and documents. In: *Proceedings of the 31st International Conference on International Conference on Machine Learning - ICML*. [S.l.: s.n.], 2014. p. II–1188–II–1196.

LIU, Q. et al. A survey on contextual embeddings. *arXiv preprint arXiv:2003.07278*, 2020

LUND, K.; BURGESS C. Producing high-dimensional semantic spaces from lexical co-occurrence. In: *Behavior research methods, instruments, & computers* 28.2 (1996): 203–208. Disponível em: <<https://doi.org/10.3758/BF03204766>>

MIKOLOV, T. et al. Efficient Estimation of Word Representations in Vector Space. *Proceedings of the First International Conference on Learning Representations*, 2013.

NOVAES, Luisa Pereira et al. Modelagem de Tópicos para a Tarefa de Recuperação de Casos Legais. *Brazilian Symposium on Databases* (2023).

RAHMAN, M. F. et al. Hdbscan: Density based clustering over location based services. *ArXiv*, abs/1602.03730, 2016.

- RUSSEL, S. J. *Artificial intelligence a modern approach*. 3. ed. Pearson Education, Inc., 2010.
- SALOMÃO, Luis Felipe (coord.). *Tecnologia aplicada à gestão dos conflitos no âmbito do Poder Judiciário Brasileiro*. Rio de Janeiro: Editora FGV Conhecimento, 2023.
- SANSONE, Carlo; SPERLÍ, Giancarlo. *Legal information retrieval systems: State-of-the-art and open issues*. Inf. Syst., Elsevier Science Ltd., GBR, v. 106, n. C, may 2022. ISSN 0306-4379. Disponível em: <<https://doi.org/10.1016/j.is.2021.101967>>.
- SCHWARCZ D. et al. *AI Tools for Lawyers: A Practical Guide*. In: *Minnesota Law Review Headnotes*. Disponível em: <<http://dx.doi.org/10.2139/ssrn.4404017>>
- SHAO, Y. et al. Bert-*pli*: Modeling paragraph-level interactions for legal case retrieval. In: *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*. [S.l.: s.n.], 2021. (IJCAI'20). ISBN 9780999241165.
- SRIVASTAVA, Akash; SUTTON, Charles. *Autoencoding Variational Inference for Topic Models*. arXiv e-prints, p. arXiv:1703.01488, mar. 2017.
- SULIS, E. et al. Exploiting co-occurrence networks for classification of implicit inter-relationships in legal texts. *Information Systems*, v. 106, p. 101821, 2022. ISSN 0306-4379. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0306437921000648>>.
- VASWANI, A. et al. Attention is all you need. In: GUYON, I. et al. (Ed.). *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2017. v. 30. Disponível em: <https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf>.
- VIANNA, D. et al. *A topic discovery approach for unsupervised organization of legal document collections*. *Artificial Intelligence and Law* (2023): 1-30.
- WALDRON, Jeremy. *The Concept and the Rule of Law*. In: No. 08-35; *Public Law & Legal Theory Research Paper Series*, Issue November, 2008.
- WIEDMANN, G. et al. Does BERT make any sense? Interpretable word sense disambiguation with contextualized embeddings. arXiv preprint arXiv:1909.10430, 2019
- YANG, J. et al. Legalgnn: Legal information enhanced graph neural network for recommendation. *ACM Trans. Inf. Syst.*, Association for Computing Machinery, New York, NY, USA, v. 40, n. 2, sep 2021. ISSN 1046-8188. Disponível em: <<https://doi.org/10.1145/3469887>>.



■ Giselle Beiguelman, frames do vídeo generativo da série Flora Rebellis, versão 2024.
Imagens geradas com Inteligência Artificial (Style Gans)

■ [Acesse o vídeo clicando na imagem](#)

