

Freepik



CRENCIAMENTO E APOIO FINANCEIRO:
PROGRAMA DE APOIO ÀS PUBLICAÇÕES CIENTÍFICAS PERIÓDICAS DA USP
COMISSÃO DE CRENCIAMENTO

Revista USP / Superintendência de Comunicação Social
da Universidade de São Paulo. – N. 1 (mar./maio 1989) -
- São Paulo, SP: Universidade de São Paulo, Superintendência
de Comunicação Social, 1989-

Trimestral.

Continuação de: Revista da Universidade de São Paulo

Descrição baseada em: N. 93 (2012)

ISSN 0103-9989

1. Ensaio acadêmico. I. Universidade de São Paulo.
Superintendência de Comunicação Social

CDD-080

dossiê inteligência artificial na pesquisa científica

5 Editorial

9 **Apresentação** *Glauco Arbix*

17 **Desafios da IA responsável na pesquisa científica** *Virgílio Almeida e Elen Nas*

29 **Inteligência artificial e mudanças climáticas** *Paulo Artaxo, Luciana Varanda Rizzo e Luiz Augusto Toledo Machado*

41 **Inteligência artificial em saúde** *Edson Amaro Jr., Helder Nakaya e Luiz Vicente Rizzo*

51 **Inteligência artificial, direito e pesquisa jurídica** *Marco Almada e Rafael A. F. Zanatta*

65 **Inteligência artificial e internet: um olhar sobre o conteúdo de usuários e a sua moderação** *Francisco Brito Cruz*

81 **Inteligência artificial, urbanização e cidades** *Márcia Regina Martins Martinez Corso, Luísa Paseto e André Carlos Ponce de Leon Ferreira de Carvalho*

91 **A inteligência artificial na pesquisa agrícola** *Antônio Mauro Saraiva, Fernando Santos Osório, André Freitas Colaço, Debora Pignatari Drucker, Eduardo Mario Mendiondo, Fernando Elias Corrêa, Filipi Miranda Soares, José Paulo Molin, Marcos Roberto Benso, Patricia Angélica Alves Marques, Roberto Fray da Silva, Sílvia Helena Galvão de Miranda, Willian França Costa e Alexandre Cláudio Botazzo Delbem*

107 **Ética em IA e IA ética: prolegômenos e estudo de casos significativos** *Luís C. Lamb*

121 **Elas na inteligência artificial – Questões de gênero** *Renata Wassermann*

133 **O futuro da pesquisa em inteligência artificial** *Anna Helena Reali Costa e Fabio Gagliardi Cozman*

textos

149 **Sociedade da mensagem para reconfigurar a des(informação)** *Magaly Prado*

165 **Temidas, rebeldes, poderosas – Milênios antes do #metoo, elas davam as cartas** *Marília Fiorillo*

livros

175 **Semeando futuros e retocando passados** *Daniel Afonso da Silva*

A **revistausp** é uma publicação trimestral da Superintendência de Comunicação Social (SCS) da USP. Os artigos encomendados pela revista têm prioridade na publicação. Artigos enviados espontaneamente poderão ser publicados caso sejam aprovados pelo Conselho Editorial. As opiniões expressas nos artigos assinados são de responsabilidade exclusiva de seus autores.

UNIVERSIDADE DE SÃO PAULO

Reitor CARLOS GILBERTO CARLOTTI JUNIOR
Vice-reitora MARIA ARMINDA DO NASCIMENTO ARRUDA

SUPERINTENDÊNCIA DE COMUNICAÇÃO SOCIAL

Superintendente EUGÊNIO BUCCI

Coordenador editorial LUIZ ROBERTO SERRANO

revistausp

Editor JURANDIR RENOVATO
Editora de arte LEONOR TESHIMA SHIROMA
Revisão MARIA ANGELA DE CONTI ORTEGA
MARIA PAULA LUCENA BONNA (estagiária)
SILVIA SANTOS VIEIRA
Secretária MARIA CATARINA LIMA DUARTE
Colaborador MARCOS SANTOS (fotografia)

Conselho Editorial
ALBÉRICO BORGES FERREIRA DA SILVA
CICERO ROMÃO RESENDE DE ARAUJO
EDUARDO VICTORIO MORETTIN
EUGÊNIO BUCCI (membro nato)
FERNANDO LUIS MEDINA MANTELATTO
FLÁVIA CAMARGO TONI
FRANCO MARIA LAJOLO
JOSÉ ANTONIO MARIN-NETO
OSCAR JOSÉ PINTO ÉBOLI

Ctp, impressão e acabamento
Gráfica CS



Rua da Praça do Relógio, 109 – Bloco L – 4º andar
CEP 05508-050 – Cidade Universitária – Butantã – São Paulo/SP
Telefax: (11) 3091-4403
www.usp.br/revistausp
e-mail: revisusp@usp.br

No início de 2020, a **Revista USP** publicou o dossiê “Inteligência Artificial”. Continha textos que, sem abrir mão da profundidade, possuíam caráter introdutório, onde se tentava explicar a um público menos habituado ao universo das inovações tecnológicas o que de fato vinha a ser a tal da IA. Naquele momento (e, convenhamos, nem faz tanto tempo assim), o que se buscava era colocar o leitor em contato com algo que, já se sabia, tinha vindo para ficar. Em seguida, irrompeu a pandemia da covid e o mundo virou de cabeça para baixo. Mas as pesquisas em IA não se mantiveram confinadas. Pouco depois estávamos falando em ChatGPT, Whisper, DALL-E etc. E, claro, nas suas consequências. Também naquela ocasião, o professor Nestor Caticha, do Instituto de Física da USP, e coordenador daquele número, lamentava o fato de ter deixado de lidar, no dossiê, com vários outros temas, e que esperava que alguns deles fossem tratados, no futuro, nesta revista.

Como se vê, não precisamos esperar tanto. O presente conjunto de artigos foi pensado e organizado por Glauco Arbix, da Faculdade de Filosofia, Letras e Ciências Humanas e do Instituto de Estudos Avançados, ambos da USP. Mais uma vez, seu trabalho à frente do dossiê foi primoroso. Segundo ele, o ano de 2023 foi “marcado pela consolidação da inteligência artificial como ponto de apoio fundamental para o avanço científico e tecnológico”. Portanto, aqui não se trata mais de desvendar os mistérios da IA, mas sim de expor e refletir sobre suas aplicações, sobretudo aquelas voltadas para a pesquisa científica, seja na área da saúde, da agricultura, do direito, da urbanização ou das mudanças climáticas. Também são discutidas as questões éticas implicadas no manejo dessas tecnologias. Tudo de forma muito clara e objetiva, sem deixar de lado os aspectos negativos e os eventuais riscos à sociedade, decorrentes de toda sorte de mau uso. Daí a necessidade, como se verá nas próximas páginas, de uma IA responsável.

Excepcionalmente, neste número, não temos a seção Arte. Em contrapartida, a concepção gráfica está toda ela alinhada com o tema do dossiê, já que as imagens utilizadas tanto nas aberturas dos artigos como na capa foram criadas por meio de inteligência artificial. Espero que gostem.

Jurandir Renovato

inteligência artificial na pesquisa científica

Apresentação

Potencial e riscos da IA na ciência

Glauco Arbix

Este dossiê apresenta um painel em que especialistas debatem como percebem e analisam o potencial da inteligência artificial (IA) na pesquisa científica, sem camuflar os riscos que representa para a vida em sociedade. A USP, mais uma vez, marca sua presença ao avaliar a mais poderosa tecnologia que a humanidade já criou.

Os embates tecnológicos e científicos ocorrem em meio ao esforço para integrar cada vez mais a IA na vida econômica e social, o que suscita dúvidas e medo sobre o grau de controle que a humanidade pode e quer ter sobre as novas tecnologias. A promessa de ganhos de produtividade e de eficiência contrasta com resultados adversos, algumas vezes inesperados, no mercado de trabalho, nas desigualdades sociais, na democracia. Dilemas éticos e morais surgem a cada experimento e jus-

tificam a busca de um marco regulatório capaz de colocar limites sem asfixiar a criatividade e a inovação.

Como se sabe, a IA não é fruto apenas de *coders*, que mergulham no mar de dados e atravessam circuitos eletrônicos, fascinados que são pelo silêncio dos laboratórios. A busca do algoritmo perdido envolve tensões econômicas e sociais, que moldam comportamentos e alteram os parâmetros dos modelos estatísticos.

A IA é a mais poderosa tecnologia que a humanidade criou. É importante demais para que apenas poucas e gigantescas corporações tenham controle praticamente total sobre suas características e recursos. No mundo todo, o poder público tem a obrigação de colocar limites e salvaguardas para o uso,

GLAUCO ARBIX é professor titular de Sociologia da FFLCH-USP, pesquisador do Center for Artificial Intelligence USP-Fapesp-IBM e do Observatório da Inovação do Instituto de Estudos Avançados (IEA) da USP e ex-presidente do Ipea e da Finep.

a pesquisa e o desenvolvimento da IA. Para manter a sociedade protegida e não apenas correr atrás do fato consumado.

O ano de 2023 foi pródigo em multiplicar iniciativas voltadas para regulamentar a IA no plano global. O G7 em Hiroshima¹, a OCDE, a atuação da *Global Partnership on Artificial Intelligence* e as Nações Unidas², assim como a cúpula em Bletchley Park, no Reino Unido³, reafirmaram que a IA pode contribuir para resolver problemas históricos da humanidade e impulsionar um crescimento econômico inclusivo e sustentável, desde que utilizada de um modo seguro, confiável, responsável e com foco no ser humano.

Embora tenham sido raras as decisões de um aparato regulatório aprovado em forma de lei, com capacidade de efetivamente colocar limites e de avançar mecanismos de proteção à sociedade, vários governos com peso na geopolítica global começaram a reconhecer os riscos que uma IA desgovernada representa para nações e povos.

As tecnologias abrigadas debaixo do termo IA aumentaram recentemente seu poder de impacto nas mais variadas dimensões da atividade humana com a dissemi-

nação de sistemas generativos, baseados em *foundation models*, a exemplo do Chat-GPT. São imensas as oportunidades abertas capazes de transformar a realidade da vida cotidiana no planeta, a começar pela educação e saúde, moradia, segurança alimentar, pesquisa científica, ambiente e clima. Os riscos, porém, são cada vez mais presentes.

A COLETÂNEA DE ARTIGOS

O presente dossiê, um exercício que vai além da informação, sugere que a reflexão crítica que permeia a vida universitária é decisiva para o avanço de uma ciência que busca o bem comum. Ainda mais após um ano como 2023, marcado pela consolidação da IA como ponto de apoio fundamental para o avanço científico e tecnológico.

De acordo com a Agência Nacional de Ciência da Austrália, 99% dos campos de pesquisa global produzem resultados com base em técnicas de IA. A ascensão dos modelos generativos, após o lançamento público de sua versão 3.5 pela OpenAI, empresa sediada na Califórnia (Estados Unidos)⁴, representou um ponto de inflexão na trajetória da IA e abriu amplas avenidas para a pesquisa no mundo todo.

De um lado, esse movimento democratiza o acesso à atividade científica. O que antes requeria *expertise* para se lidar com a linguagem dos programadores, agora é feito com ferramentas amigáveis e de fácil

1 *G7 Action Plan for promoting global interoperability between tools for trustworthy AI* (Hiroshima, 20/5/2023).

2 O secretário-geral da ONU criou um *AI Advisory Body on risks, opportunities, and international governance of artificial intelligence*, com mais de 30 especialistas de alto nível dos mais diferentes países (Nova York, 27/10/2023).

3 28 países e a União Europeia se reuniram no Reino Unido e assinaram a *The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023*. Assinantes: Austrália, Brasil, Canadá, Chile, China, União Europeia, França, Alemanha, Índia, Indonésia, Irlanda, Israel, Itália, Japão, Quênia, Arábia Saudita, Holanda, Nigéria, Filipinas, Coreia, Ruanda, Singapura, Espanha, Suíça, Turquia, Ucrânia, EAU, Reino Unido e EUA.

4 A OpenAI foi criada em 2015 como entidade de pesquisa sem fins lucrativos. Em 2019 desenvolveu seu braço empresarial (com fins lucrativos), tornando-se uma instituição híbrida, que recebeu volumosos aportes de investimento da Microsoft.

acesso, com impactos positivos no universo do conhecimento e da sociedade. De outro, gera novos problemas e arrisca piorar o que já não era bom, como os empregos, a desigualdade e a própria democracia.

Mesmo com esses senões de luzes e sombras, a IA vem sendo abraçada por cientistas de todas as linhagens e tem se mostrado capaz de transformar o fazer científico e a atividade própria dos cientistas.

Para alguns expoentes da IA, como Demis Hassabis, neurocientista e fundador da DeepMind⁵, “estamos no renascimento da descoberta científica [...] e a IA tem um potencial incrível para ajudar a resolver os maiores desafios da humanidade”. A IA é um “multiplicador da engenhosidade humana”, completou Hassabis, que permite compreender e enxergar o que os cientistas não conseguem a olho nu⁶. Grande número de cientistas parece concordar com Hassabis e reconhecer o papel crescente que a IA desempenha nos mais diferentes campos da ciência.

Entre 1.600 cientistas entrevistados pela revista *Nature*⁷, mesmo sem deixar de registrar suas preocupações com os pontos negativos, cerca de 55% relataram se beneficiar da IA generativa para editar e traduzir textos do inglês; 40% utilizaram a IA para

acelerar e programar; mais de 30% para redigir artigos e para ampliar a pesquisa; 25% para obter ideias criativas; 15% para gerar novas hipóteses. A pesquisa registrou também que 58% dos cientistas consideram que a IA produz resultados não realizáveis por ferramentas tradicionais.

Para o Conselho Europeu de Pesquisa, em seu relatório de dezembro de 2023, 85% dos pesquisadores afirmaram que a IA generativa poderia assumir tarefas rotineiras e revisões de literatura, sendo que 40% consideraram que a IA aumenta significativamente a produtividade na ciência⁸.

O AVANÇO DA IA NA CIÊNCIA

A intensa penetração da IA nos sistemas de pesquisa e inovação pode ser percebida por alguns dos principais avanços que marcaram a ciência no ano de 2023 e que dificilmente seriam alcançados sem o concurso de tecnologias de IA.

No último dezembro, a FDA⁹ aprovou o primeiro e único medicamento baseado na tecnologia CRISPR¹⁰, o Casgevy, desenvolvido pela Vertex Pharmaceuticals (de Boston) e pela suíça CRISPR Therapeutics, capaz de tratar duas doenças: a anemia falciforme (doença sanguínea crô-

5 Empresa sediada em Londres cuja contribuição foi chave para a formulação das técnicas de aprendizagem profunda (*deep learning* – DL), baseada em redes neurais, que está na base da expansão da IA a partir de 2010. Ainda que a IA faça parte da caixa de ferramentas dos cientistas desde a década de 60, foi a partir da DL que abraçou o mundo.

6 Demis Hassabis, em palestra na University College London, em novembro de 2023, quando retomou a visão que tem sobre a IA desde entrevista à *The Economist*, em 2020.

7 Richard Van Noorden; Jeffrey M. Perkel. “AI and science: what 1,600 researchers think”. *Nature*, v. 621, 28/sep./2023.

8 *Foresight: use and impact of AI in the scientific process*. European Research Council, dec./2023.

9 Food and Drug Administration (FDA) é a agência pública norte-americana que controla alimentos, medicamentos, materiais biológicos e equipamentos de modo a garantir sua eficácia e segurança.

10 Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) é uma espécie de tesoura genética que permite cortes no código genético de uma célula, a partir da edição de seu DNA.

nica que afeta cerca de mais de 60 mil pessoas somente no Brasil, basicamente a população negra) e a β -talassemia (também genética e hereditária). Mais de 70% dos portadores de anemia falciforme vivem na África subsaariana.

Ainda no campo da saúde, a *Science* destacou o Ozempic e o Wegovy (nomes de mercado), produzidos pela Novo Nordisk (Dinamarca), para tratamento inovador de diabetes e obesidade e que também foram liberados pela FDA e autoridades europeias¹¹.

Também foram dignos de nota no ano os avanços promovidos pelo sistema AlphaFold2, desenvolvido pela Google DeepMind, que previu em 3D mais de 200 milhões de estruturas de proteínas e abriu espaço para a descoberta de novas drogas¹². E, na mesma direção, o destaque coube à primeira vacina contra a malária com mais de 80% de eficácia, desenvolvida pela Universidade de Oxford. Claro que avanços de ruptura sempre devem ser tratados com a devida cautela. A abordagem de novos métodos e medicamentos, assim como a discussão sobre a importância das novas tecnologias, fazem parte das preocupações do artigo “Inteligência artificial em saúde”, neste dossiê.

Novos instrumentos e metodologias para combater as mudanças do clima compõem os avanços científicos que ganharam destaque no ano passado. Ao contribuir para o enfrentamento da maior ameaça que atinge

a humanidade, a IA estimula a reação de todos os países para mitigar as consequências negativas das alterações climáticas. No médio e longo prazos, os modelos analíticos com base em algoritmos cada vez mais complexos serão cada vez mais fundamentais para as previsões do tempo e do clima. Os ganhos de precisão em modelos como os *earth system models* (ESM) permitem analisar as interações entre a atmosfera, oceano, ecossistemas terrestres, gelo e biosfera, como registrado no instigante artigo “Inteligência artificial e mudanças climáticas”.

No mesmo sentido, a busca de novas fontes de energia, de materiais avançados e novos experimentos se multiplica e anima redes globais de pesquisadores. No planejamento urbano, os profissionais de IA são cada vez mais requisitados para dar conta da complexidade da segurança e da mobilidade nas cidades do século XXI, que precisam se tornar mais inteligentes e sustentáveis, como aponta o artigo “Inteligência artificial, urbanização e cidades”, para melhorar a qualidade de vida de suas populações.

Na agricultura, uma das áreas mais importantes da economia brasileira, a IA está nos alimentos, nas fibras e na energia que movimenta empregos e racionaliza o uso do solo e da água, ao mesmo tempo em que ergue barreiras de defesa do meio ambiente. O artigo “A inteligência artificial na pesquisa agrícola” ilumina os avanços e os principais desafios que a agricultura enfrenta e que não podem prescindir das contribuições da IA.

UM ECOSISTEMA PARA A IA

É preciso ir muito além das aplicações do conhecimento. Visão e governança precisam

11 Essas novas terapias não somente alteram como a obesidade e a diabetes são tratadas, mas também como são compreendidas. Porém, como toda abordagem de ruptura, outros efeitos são sentidos, nem sempre benéficos (*Science*, v. 382, issue 6676, 14/dec./2023).

12 Ver Database Protein Structure. Disponível em: <https://alphafold.ebi.ac.uk>.

ser reforçadas com a urgente formação de profissionais de IA e a construção de um sistema de dados capaz de valorizar os grandes ativos sediados no SUS, Inep, IBGE e tantos outros órgãos. A valorização de um ecossistema de IA consistente e diversificado ajuda a superar obstáculos crônicos (como o subfinanciamento) e a estimular a formação de uma comunidade de pesquisadores capaz de desenvolver uma IA inclusiva e sintonizada com a resolução dos problemas mais sentidos pela população brasileira.

Apenas um ecossistema nacional para IA, com suas regras e instituições, poderá abrir caminho para o equacionamento de questões éticas – como as dificuldades para explicar resultados científicos – ou para atuar no aumento da diversidade na pesquisa, que continua sendo marcada como um mundo masculino e branco. Passos dessa natureza ajudam a elevar a confiança na ciência e a preparar uma nova geração de pesquisadores, mais cosmopolitas, capaz de refletir sobre problemas locais, nacionais e globais. Sem abrir essa oportunidade, será difícil para o Brasil participar da criação de instituições globais de governança, aptas a definir os limites de uma IA e, assim, ajudar na construção da paz. Esse horizonte atrai jovens pesquisadores e ajuda a universidade a quebrar a atual estagnação do conhecimento que limita sua expansão.

IMPACTOS SOCIAIS DA IA

O Brasil não está na linha de frente da produção e desenvolvimento da IA. Como a maior parte dos países, o Brasil também tem dificuldades de infraestrutura computacional, de semicondutores, de acesso aos

bancos de dados que orientam os grandes modelos de linguagem existentes hoje. Dominados por poucas empresas, vivem o simulacro da autorregulação, que não tem efetividade e apenas prolonga a política que comandou a digitalização desde os anos 70 e que resultou no descontrole e na fragmentação atual, em que as sociedades são fustigadas pela desinformação massiva, que mina a busca da veracidade dos fatos, substrato básico da democracia.

Atualmente, tendências distintas estão em confronto e não contam com solução à vista. Uma destas caminha na direção de maior concentração e escala, procurando desenvolver sistemas cada vez maiores e mais poderosos. Ou seja, identificam melhor desempenho com maior poder computacional, mais dados e mais parâmetros. É o que movimenta os grandes modelos e responde por investimentos na casa de bilhões de dólares, difíceis de serem acompanhados pela maioria dos países e, claramente, por pequenas empresas. Por outro lado, sistemas de código aberto e menores procuram se aproximar das práticas de fronteira.

Seja qual for o desenlace desse confronto, a multimodalidade avançou rapidamente em 2023; ou seja, a capacidade da IA processar diferentes tipos de dados, não apenas textos, mas também imagens, vídeos, áudios e sons. Isso significa que os grandes modelos podem ser treinados em novos e vastos conjuntos de dados, oriundos de diferentes fontes, e que superam em muito as informações contidas apenas em textos. Essa tendência nem sempre ajuda os alinhamentos de valores, a possibilidade de sistemas de IA responderem em sintonia com os valores humanos, que, antes de tudo, são diferenciados e mutantes.

Esses processos procuram evitar respostas racistas e sexistas dos grandes modelos com

trabalho humano, com base em técnicas de *reinforcement learning*. Ou seja, avaliadores humanos identificam respostas inadequadas e treinam o sistema sobre o que é ou não aceitável. Como as variáveis e modulações são muitas, as falhas continuam a fazer parte do cotidiano dos modelos generativos, mesmo com todo o esforço de aprimoramento.

Pesquisadores do Allen Institute¹³ mostraram recentemente que a tentativa de aumentar o volume de dados e de utilizar o *reinforcement learning* para diminuir vieses pode resultar praticamente em seu contrário. A pesquisa testou cinco *chatbots* entre os mais populares, que responderam de modo diferente quando a pergunta no *prompt* era feita com a linguagem dos negros americanos ou em inglês padrão. Essa sensibilidade à linguagem diferente do inglês tido como correto expressou, segundo os pesquisadores, um racismo encoberto. Na verdade, os preconceitos que permeiam os bancos de dados e os algoritmos indicam a falta de diversidade no ambiente de pesquisa e nos grupos que desenvolvem a IA, como se pode acompanhar pela leitura do artigo “Elas na inteligência artificial – Questões de gênero”, tema que alimenta o debate sobre o florescimento de questões éticas, trabalhadas neste dossiê em um texto elegantemente construído com o título “Ética em IA e IA ética: prolegômenos e estudo de casos significativos”. O debate crescente em torno desse tópico apenas demonstra que as tecnologias disruptivas geram impactos que vão muito além das atividades de sua base tecnológica. No mesmo sentido, a circulação

intensa de conteúdos pelos canais da internet apenas reforça que as distorções da IA podem perpetuar vieses e injustiças sistêmicas. Como os dados e informações da rede global sustentam grande parte dos modelos generativos, o entrelaçamento do debate ético com a moderação de conteúdos torna-se inevitável. O trabalho incluído nesta coletânea “Inteligência artificial e internet: um olhar sobre o conteúdo de usuários e a sua moderação” ajuda a visualizar os riscos e o esforço de proteção que as plataformas de internet e os grandes modelos de linguagem precisam enfrentar. Essa perspectiva é ampliada pelo texto “Inteligência artificial, direito e pesquisa jurídica”, que procura mapear as implicações da institucionalização de diferentes projetos de IA no sistema de justiça brasileiro e as consequências para a pesquisa em direito.

A IA ENTRA NO DNA DA CIÊNCIA

A presente coletânea de artigos oferece uma pequena amostra de como a IA se tornou relevante e foi abraçada no mundo todo pela ciência e pelos pesquisadores, em um movimento que parece irreversível. Globalmente, o número de artigos em revistas científicas triplicou em uma década, de 2010 a 2021, enquanto os artigos baseados em *machine learning* quadruplicaram no mesmo período¹⁴.

A imprecisão, a segurança, as alucinações persistem como grandes obstáculos ao rápido crescimento da IA generativa nas economias. Mesmo assim, os dados são inequívocos: a partir de 2016, a IA saltou dos bancos

13 V. Hofmann et al. “Dialect prejudice predicts AI decisions about people’s character, employability, and criminality”, 2024. Disponível em: <https://arxiv.org/abs/2403.00742>.

14 Center for Security and Emerging Technology Report, 2022.

acadêmicos para os laboratórios de grandes corporações, que lideram desde então a geração de modelos generativos. Como *players* globais, mostram-se mais fortes que a maioria dos países. O controle que exercem sobre gigantescos bancos de dados e um parque de computadores de alto desempenho são uma amostra do nível de investimento que realizam para se reafirmarem como agentes globais determinantes do presente e provavelmente do futuro da IA. Esses grandes conglomerados expandem a IA ao mesmo tempo que constroem governos e inibem o esforço de construção de sistemas reguladores globais e nacionais.

Mais um paradoxo do mundo de hoje, em que o número de incidentes e controvérsias provocados pela IA cresceu 26 vezes entre 2012 e 2021, segundo o relatório da AI Index deste ano, publicado pela Universidade de Stanford. A opacidade das ferramentas da IA e suas consequências negativas, como a discriminação, as falsidades e a desinformação, são tratadas neste dossiê no texto “Desafios da IA responsável na pesquisa científica”.

O DECLÍNIO DA PESQUISA DE RUPTURA

A ciência precisa de uma IA transparente, capaz de ir além da lógica que preside a pesquisa há décadas, e que se apoia na revisão por pares, nos laboratórios universitários e nas declarações em favor da interdisciplinaridade.

O esforço de superação se justifica porque estudos recentes mostram uma situação incômoda para a ciência, ao apontar um arrefecimento do progresso da pesquisa nos grandes campos científicos. Nas palavras dos

pesquisadores, o número de artigos cresceu, mas não o de *papers* de ruptura, aqueles que abrem novas fronteiras para o conhecimento. Esse estudo foi publicado pela *Nature*, em 2023, e cobriu 45 milhões de artigos e 39 milhões de patentes, rastreados nos últimos 60 anos. Os resultados são fortes e pedem muita reflexão sobre as hipóteses que tentam explicar essa queda.

É possível que a ciência esteja com dificuldades para lidar com o enorme aumento do conhecimento acumulado; por isso, tenderia a apanhar os *low-hanging fruits*, as frutas mais maduras, os *insights* com origem nos principais autores e nas revistas mais importantes, tentando com isso trilhar os caminhos da pesquisa de menor risco. Uma segunda hipótese é que o crescimento da pesquisa não foi acompanhado do devido financiamento, ou seja, o subfinanciamento da ciência dificultaria os artigos de ruptura. A terceira hipótese, corrente nas universidades, é que os cientistas, pressionados e ávidos por publicar, tenderiam a diminuir a experimentação e sua ousadia.

A quarta hipótese, a que mais tem aderência à IA, sugere uma quase exaustão de recursos, metodologias e de teorias explicativas para novos fenômenos. O aprofundamento desta hipótese indica que o declínio das descobertas disruptivas pede mudanças fundamentais na natureza da produção científica e tecnológica.

A IA, claramente, tem potencial para ajudar em todas essas dimensões.

A TRAVESSIA ATUAL

Exatamente por isso, a IA se transformou em objeto de desejo de países,

empresas e universidades, que querem alavancar sua presença na arena mundial, nos mercados e nos domínios do saber. Em outras palavras, a IA passou a ocupar uma posição insubstituível para impulsionar todas as fases da pesquisa, seja na assistência, na automação, na previsão e na construção de cenários e estratégias de produção de conhecimento novo.

Apesar de amplamente disseminada pelo planeta, os destinos da inteligência artificial ainda não são nítidos. Todas as tensões presentes em sua criação, desenvolvimento e uso relatadas neste dossiê são expressão de uma disputa aberta entre países, empresas, universidades e diferentes grupos sociais. São dissonâncias de interesse, de hábito, de história, que diferenciam as intenções do gesto efetivo.

É importante lembrar que a IA está em sua primeira infância e suas características ainda estão em plena formação. Suas configurações aguardam definição, testes e provas de conceito. Ao se olhar para um futuro mais distante, combinações entre métodos simbólicos e baseados em dados provavelmente receberão maior atenção, enquanto apostas como computação quântica poderão levar a novos saltos de qualidade, como aponta o artigo desta coletânea “O futuro da pesquisa em inteligência artificial”.

Nossa época é de trânsito. E as novas tecnologias estão envelopadas pelo social, não o contrário. A proatividade é parte da ciência moderna, que se destacou como uma construtora de mundos, que recusou a passividade destilada pelas diferentes versões do determinismo tecnológico.

A ciência ousada recusa o deslumbremento e o medo diante do novo e do desconhecido. Por isso mesmo, o abraço sem a crítica e a crítica impotente da IA não conseguem explicar e muito menos modificar o curso, a configuração e os propósitos dessa tecnologia que altera padrões e sacode instituições e suas regras, a começar pela universidade e seus mecanismos de ensino, aprendizagem e pesquisa.

A certeza que se pode ter neste momento, diante de impactos diferenciados que a IA provoca na sociedade, é que nossas vidas serão muito diferentes. Em alguns domínios, como na ciência, a reviravolta deverá ser grande e duradoura, o que levará pesquisadores a abordar a IA com a prudência apropriada aos tempos de hoje. É o que está registrado neste dossiê: um tratamento tão atual quanto responsável da IA, temperado pela recomendação sensata do físico Mario Krenn, do Max Planck Institute: “Aproximem-se da IA como de uma musa, uma fonte de inspiração e de ideias”¹⁵.

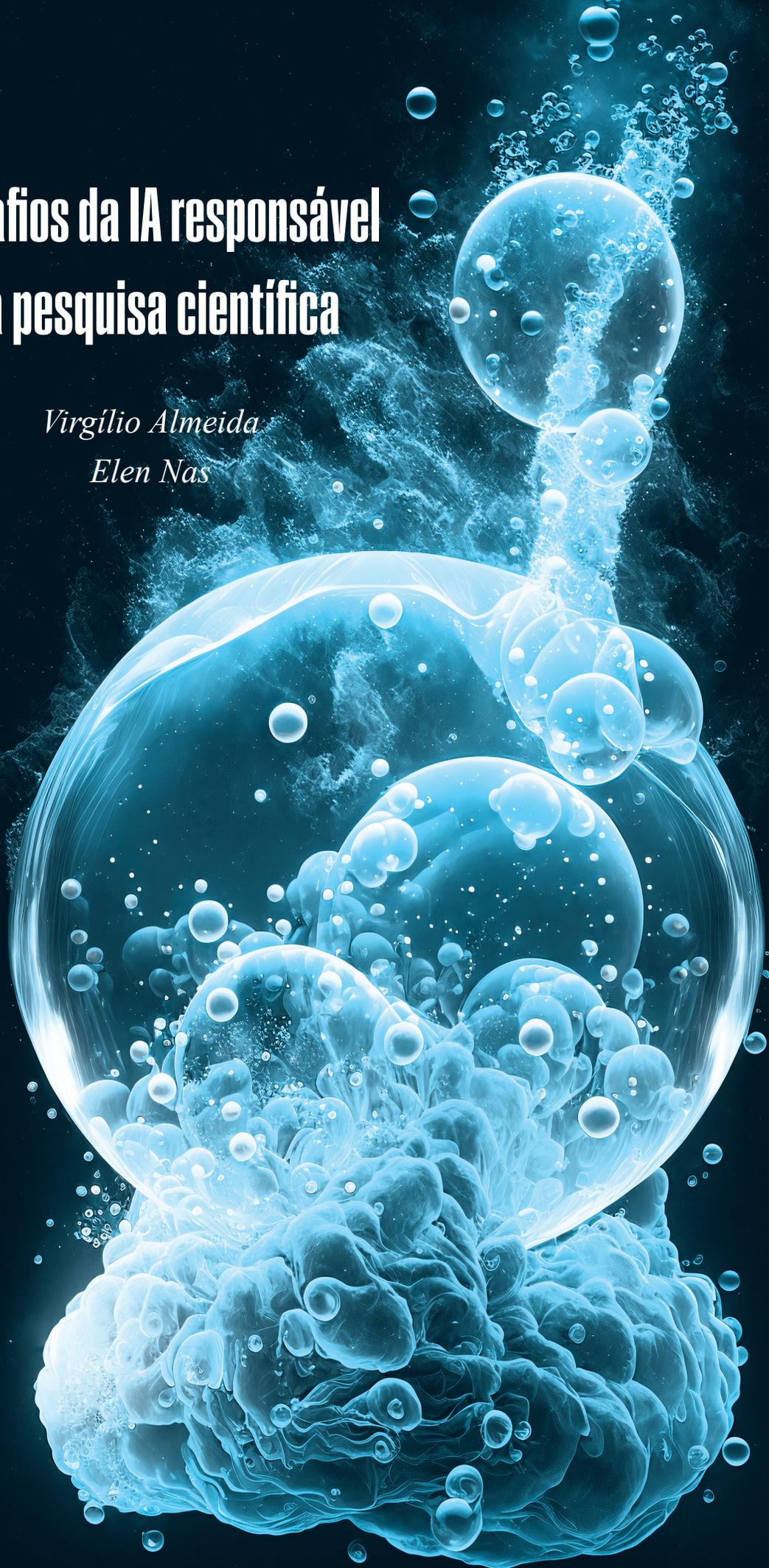
Boa leitura.

15 Ver: <https://mariokrenn.wordpress.com/>.

Desafios da IA responsável na pesquisa científica

Virgílio Almeida

Elen Nas



resumo

A inteligência artificial está cada vez mais presente em praticamente todas as áreas da ciência e tecnologia. O uso de tecnologias de IA tem o potencial de acelerar a produtividade dos pesquisadores e reduzir custos da pesquisa científica. Apesar dos benefícios, preocupações surgem em relação à opacidade das ferramentas de IA, que podem levar a consequências negativas, como discriminação, falsidades e desinformação científica, comprometendo a confiança pública na ciência. O artigo explora a implementação dos princípios de IA responsável nas organizações de pesquisa visando minimizar possíveis impactos negativos da IA na ciência.

Palavras-chave: IA responsável; transparência; inteligência artificial; IA ética; centralidade do humano.

abstract

Artificial intelligence is gaining prominence across every sector of science and technology. Leveraging AI technologies holds promise for boosting researchers' productivity and cutting down on scientific research costs. However, concerns have emerged regarding the lack of transparency in AI tools, which may result in adverse outcomes like discrimination, scientific misinformation, and eroded public trust in science. This article delves into how research organizations are adopting responsible AI principles to mitigate potential negative effects on science.

Keywords: *responsible AI; artificial intelligence; transparency; ethical AI; centrality of the human.*

O

destino das tecnologias de inteligência artificial não está previamente determinado; há um elevado grau de incerteza e imprevisibilidade em relação ao seu uso e desenvolvimento. O futuro está em aberto e depende em boa medida do que o país for capaz de formular e construir. É essencial compreender o potencial de utilização e progresso da IA no Brasil, especialmente considerando a desigualdade econômica e o baixo nível de escolaridade, que representam desafios para a implementação das novas tecnologias no país. Ao analisar o papel da IA no contexto brasileiro, é preciso entender como essas tecnologias afetam a produtividade, o mercado de trabalho

e a inovação. Mas é também fundamental avaliar os riscos e as consequências da IA em países em desenvolvimento, como o Brasil.

Inteligência artificial, com seus sistemas algorítmicos e robôs, tem sido usada de muitas maneiras socialmente benéficas, como, por exemplo, antecipar necessidades de saúde, auxiliar gerência de tráfego em centros urbanos, otimizar o uso de recursos escassos, facilitar decisões financeiras e aumentar a eficiência de serviços governamentais. Algoritmos têm sido também usados para tomar decisões críticas

VIRGÍLIO ALMEIDA é professor emérito do Departamento de Ciência da Computação da UFMG e professor associado ao Berkman Klein Center da Universidade de Harvard.

ELEN NAS é pesquisadora da Cátedra Oscar Sala do Instituto de Estudos Avançados (IEA) da USP.

para indivíduos, com impacto direto na vida das pessoas. Baseados em grandes massas de dados, algoritmos operam as plataformas de mídia social, controlando o fluxo de informações para diferentes setores da sociedade, com impactos na formação social, cultural e política dos indivíduos e no comportamento coletivo da sociedade. Spotify, Twitter, YouTube, Waze, Netflix, Facebook, TikTok, Amazon são algumas das plataformas algorítmicas que impactam bilhões de vidas em todo o mundo, influenciando hábitos, culturas e valores compartilhados por diferentes grupos sociais. Há, no entanto, uma preocupação crescente em relação ao papel dos algoritmos na sociedade, em particular com a tomada de decisão algorítmica sobre indivíduos e grupos sociais. Por exemplo, algoritmos projetados para filtrar, selecionar e exibir a grande quantidade de informações disponíveis online, combinados com a tendência das pessoas de buscar ambientes sociais similares, podem induzir a vieses em grupos da sociedade, alimentando racismo, preconceitos, discriminação e desinformação. Inteligência artificial, robôs e demais máquinas operadas por algoritmos mediam cada vez mais nossas interações profissionais, sociais, culturais, econômicas e políticas. Compreender o comportamento dos algoritmos e sistemas de inteligência artificial é essencial para nossa capacidade de governar suas ações, colher benefícios e minimizar seus impactos negativos (Mendonça; Filgueiras; Almeida, 2023).

As tecnologias de inteligência artificial já afetam vários aspectos da ciência e tecnologia. São tecnologias que podem

analisar grandes volumes de dados com o objetivo de classificar, categorizar, identificar padrões, prever e tomar decisões baseadas em evidências. O desenvolvimento e aplicação da IA ocorre em diferentes aspectos das ciências, incluindo biologia, física, matemática e ciências sociais. O uso das tecnologias de IA na ciência, como aprendizado de máquina, aprendizado profundo, processamento de linguagem natural e tecnologias de IA generativa tem o potencial de acelerar a produtividade de pesquisadores, reduzir custos de experimentos e levar qualidade e segurança à pesquisa científica. Por exemplo, formular e testar hipóteses são etapas fundamentais do método científico, que demandam tempo e conhecimento nas pesquisas científicas. Os métodos de IA podem contribuir em várias fases do processo de geração de hipóteses, a partir da identificação de padrões e expressões simbólicas em grandes massas de dados experimentais. Essas tecnologias de IA são capazes de acelerar a identificação de contraexemplos que contradizem uma conjectura teórica (Wang, 2023). Há a expectativa de que a IA possa levar a um aumento de produtividade necessária para a ciência avançar na compreensão dos grandes desafios globais da humanidade, como mudanças climáticas, poluição, escassez de recursos e doenças infecciosas.

Os avanços da inteligência artificial na pesquisa são evidenciados pelo aumento de publicações científicas que mencionam IA nos títulos e resumos. Um levantamento da revista *Nature* mostra que a proporção de artigos na base de dados Scopus mencionando termos referentes

à IA nos títulos ou resumos agora é de 8%, um aumento de 2% em relação a uma década atrás. A pesquisa da *Nature* coletou opiniões sobre o papel da IA ao entrevistar mais de 1.600 pesquisadores de todo o mundo. Os pesquisadores identificaram várias formas positivas de uso das ferramentas de IA na pesquisa científica. Dois terços dos entrevistados observaram que a IA acelera o processamento de grandes massas de dados, 58% informaram que a IA permite a execução de cálculos que anteriormente não eram viáveis e 55% mencionaram economia de tempo e recursos nos projetos de pesquisa (Van Noorden, 2023).

Ao mesmo tempo que a maioria dos pesquisadores mostrou aspectos positivos no uso de IA na pesquisa científica, foram também identificadas preocupações com essas tecnologias, que incluem principalmente a opacidade das ferramentas de IA, consideradas por muitos “caixas-pretas”, das quais não é possível entender as razões subjacentes aos resultados gerados por IA. As preocupações também abrangem a existência de vieses, bem como as possibilidades de estudos com resultados falsos e propagação de desinformação científica. A consequência desses aspectos potencialmente negativos é o comprometimento da confiança pública na ciência e nos resultados científicos.

Este artigo explora as possibilidades de se adequar os fundamentos e práticas da inteligência artificial responsável (IAR) ao contexto das organizações de pesquisa, principalmente dentro de universidades, pois no Brasil mais de 95% da produção científica nas bases internacionais deve-se à capacidade de pes-

quisa de universidades públicas, federais e estaduais (Moura, 2019).

ENTENDENDO A IA RESPONSÁVEL

A IA responsável visa ao desenvolvimento de sistemas algorítmicos que beneficiem indivíduos, sociedades e o meio ambiente. Abrange os aspectos éticos, legais e técnicos do desenvolvimento e implementação de tecnologias de IA benéficas. Isso inclui garantir que o sistema de IA não cause danos, não interfira na agência humana, não discrimine e não desperdice recursos. Assim, a IA responsável é uma área de pesquisa multidisciplinar, em que cientistas da computação, éticos, juristas, sociólogos, psicólogos e muitos outros especialistas trabalham juntos para moldar o futuro da tecnologia, visando ao bem-estar humano como principal objetivo (Baeza-Yates, 2022).

No ambiente de pesquisa científica, há necessidade de formular estratégias visando apoiar as equipes de pesquisa científica para o desenvolvimento e implementação de IA responsável, com fundamentos éticos, responsabilidades morais e sociais, que estabeleçam limites de proteção para sociedade e meio ambiente. As organizações de pesquisa científica que estão envolvidas no desenvolvimento, implementação e uso de ferramentas de IA têm observado potenciais conflitos entre avanço tecnológico e preocupações éticas e sociais com o impacto dessas novas tecnologias. A tendência em direção à IA responsável nessas instituições destaca um crescente foco na incorporação de considerações éticas no processo de *design* e

desenvolvimento de ferramentas de IA, promovendo valores fundamentais como *responsibilities*, transparência, justiça e inclusão social.

Os fundamentos básicos da IA responsável buscam garantir que os sistemas de IA sejam seguros, aceitos e confiáveis. Para isso, o sistema deve ser projetado levando em consideração princípios éticos e as consequências morais de suas ações e decisões, de maneira responsável e transparente (Dignum, 2017). Diversas organizações internacionais têm elaborado conjuntos de princípios fundamentais para a adoção responsável da inteligência artificial. Estes princípios visam mitigar os riscos específicos associados ao uso dessas tecnologias, tais como equidade, viés, autonomia, privacidade, proporcionalidade e responsabilidade. A minimização dos riscos da inteligência artificial requer o estabelecimento de normas e princípios relacionados à transparência, interpretabilidade, explicabilidade e auditabilidade, entre outros aspectos. Tais princípios servem como diretrizes para orientar o desenvolvimento de aplicações de inteligência artificial em diversos domínios, como ciência, saúde, educação, finanças e mídias sociais, apenas para mencionar algumas áreas impactadas por essa tecnologia. É fundamental que os princípios de inteligência artificial responsável sejam complementados por uma governança adequada dos dados e modelos de IA, além do treinamento das pessoas capacitadas para sua implementação.

O Conselho de Políticas Tecnológicas da ACM (Association for Computing Machinery) elaborou uma Declaração de Princípios para Sistemas Algorítmicos

Responsáveis (Council, 2022), dado que os algoritmos de IA estão sendo cada vez mais utilizados por empresas, governos e cidadãos, com impactos significativos sobre indivíduos e sociedade. O documento da ACM apresenta um conjunto de princípios fundamentais para a criação de sistemas de IA responsáveis. Esse conjunto abrange os seguintes princípios: 1) legitimidade e competência; 2) minimização de danos; 3) segurança e privacidade; 4) transparência; 5) interpretabilidade e explicabilidade; 6) manutenibilidade; 7) contestabilidade e auditabilidade; 8) responsabilidade (*accountability*); e 9) limitação de impactos ambientais (Council, 2022).

Implementar diretrizes de IA responsável e ética requer transformar princípios abstratos em medidas concretas, procedimentos e práticas que possam ser incorporados aos sistemas, ferramentas e fluxos de trabalho de uma organização. Uma implementação eficiente dos princípios conduz a uma maior transparência e confiança entre os usuários e as partes interessadas, diminuindo a distância entre os conceitos teóricos e sua aplicação no mundo real.

Os países da Organização para a Cooperação e Desenvolvimento Econômico (OCDE) adotaram a Recomendação do Conselho sobre Inteligência Artificial. Os princípios da OCDE sobre Inteligência Artificial promovem uma IA inovadora, confiável e que respeite os direitos humanos e os valores democráticos. Esses princípios se aplicam aos atores de IA e a todo o governo, incluindo a implementação de IA (Yeung, 2020).

A OCDE estabeleceu os seguintes princípios para a administração responsável de IA confiável:

- crescimento inclusivo, desenvolvimento sustentável e bem-estar, visando engajar os agentes proativamente na busca responsável de uma IA confiável em busca de resultados benéficos para as pessoas e o planeta;
- valores centrados no ser humano e justiça, que devem respeitar os direitos humanos e os valores democráticos ao longo do ciclo de vida dos sistemas de IA;
- transparência e explicabilidade, que buscam oferecer uma compreensão geral dos sistemas de IA, para permitir que aqueles afetados por um sistema de IA entendam os resultados que lhes afetam;
- robustez e segurança, que devem garantir sistemas de IA seguros em relação aos dados e decisões tomadas por esses sistemas;
- responsabilidade, visando tornar os agentes de IA responsáveis pelo funcionamento adequado dos sistemas de IA.

Esses princípios da OCDE ressaltam pontos que são comuns a vários outros documentos sobre IA responsável, que são os pilares fundamentais: transparência, responsabilização, explicabilidade e justiça. Além dos princípios, há também ênfase na importância de monitoramento permanente das aplicações de IA e na necessidade de investimentos públicos e privados em pesquisa e desenvolvimento para estimular a inovação em IA responsável.

Cada vez mais, IA está sendo implantada em ambientes de ensino superior para

aprimorar a eficiência dos sistemas, refinar tomadas de decisão e melhorar a prestação de serviços. As universidades estão começando a adotar ferramentas habilitadas por IA para tarefas que variam desde administração até ensino e pesquisa. No entanto, o uso de IA pode acarretar riscos significativos. A Universidade da Califórnia (UC), como uma grande universidade pública, desenvolveu uma estratégia para orientar a implantação das tecnologias de IA, de maneira a proteger a comunidade e refletir os valores da instituição, que incluem não discriminação, segurança e privacidade. Assim, a UC adotou um conjunto de recomendações para orientar a implantação segura e responsável da inteligência artificial e estabeleceu os seguintes princípios para aquisição, desenvolvimento, implementação e monitoramento de ferramentas de IA dentro dos diversos *campi* da universidade (University of California, 2021):

- avaliação cuidadosa dos benefícios e riscos da IA, considerando que as necessidades das pessoas afetadas são essenciais para decidir sobre sua aplicação;
- transparência é fundamental; as pessoas devem ser informadas sobre o uso de ferramentas de IA e capazes de entender os resultados e contestá-los, se necessário;
- as ferramentas de IA devem ser eficazes, precisas e seguras para seu propósito, ao longo de sua vida útil;
- é crucial avaliar e mitigar preconceitos e discriminação nas ferramentas de IA, adotando procedimentos para remediar danos;
- privacidade e segurança dos dados devem ser priorizadas no *design* das ferramentas habilitadas por IA;

- desenvolvimento e uso de IA devem respeitar valores humanos e direitos civis e humanos, observando as leis e princípios aplicáveis;
- as ferramentas de IA devem promover benefícios equitativos para todos, incluindo aspectos sociais, econômicos e ambientais;
- a Universidade da Califórnia deve ser responsável pelo desenvolvimento e uso ético de sistemas de IA em seus serviços, de acordo com os princípios acima.

GOVERNANÇA DE IA RESPONSÁVEL

A existência de conjuntos de princípios destinados a orientar pesquisadores e professores no desenvolvimento e na implementação de tecnologias de IA é um ponto de partida útil, porém insuficiente, pois não estão associados a explicações sobre como os princípios poderiam ser implementados na prática. A questão fundamental a ser explorada é: que tipos de estruturas de governança precisam estar em vigor para garantir o uso responsável de IA na pesquisa científica?

A IA responsável ajudará a minimizar consequências indesejáveis e não intencionais, como danos à sociedade, ao meio ambiente e à confiança que a sociedade tem nos resultados científicos. À medida que o uso de IA permeia uma parcela maior de laboratórios e organizações de pesquisa, é necessário gerenciar o risco do uso dessas tecnologias emergentes. Os princípios são importantes, mas só proporcionarão o que as organizações de pesquisa precisam se forem combinados com práticas de governança que ajudem

a guiar o uso de ferramentas de IA em projetos de pesquisa científica, desde o *design* até a produção e publicação de resultados de pesquisa.

Um estudo publicado pelas Academias Nacionais dos Estados Unidos, que inclui as academias de Medicina, Engenharia e Ciências, explora questões éticas e de governança relacionadas às consequências individuais e sociais da pesquisa em computação e suas aplicações, que incluem inteligência artificial (National Academies of Sciences, 2022). Quando os resultados da pesquisa em computação, como IA, são incorporados em sistemas com potencial impacto social, várias dimensões da responsabilidade social precisam ser abordadas por meio de mecanismos de governança e por órgãos reguladores.

Para evoluir em direção ao uso responsável de IA, as instituições, sejam elas laboratórios ou universidades, devem fornecer orientações sobre estratégias para identificar e avaliar os riscos de ferramentas habilitadas por IA e estabelecer mecanismos de governança apropriados para a instituição se adequar às políticas e normas para uso de IA (Bano, 2023).

TREINAMENTO

É essencial avaliar a compreensão dos conceitos de ética e responsabilidade em IA dentro de uma organização de pesquisa, para orientar estratégias personalizadas, modelos de governança e programas de treinamento adaptados às necessidades específicas do grupo de pesquisadores. Um estudo conduzido junto a pesquisadores do CSIRO (Commonwealth Scientific

and Industrial Research Organization), na Austrália, oferece informações importantes sobre a percepção de pesquisadores e professores referentes a ética e responsabilidades em projetos de pesquisa que fazem uso das tecnologias de IA. Os projetos avaliados, usando tecnologias de IA, situam-se em vários domínios da ciência.

Os resultados apontam para a necessidade de maior conscientização e apoio para abordar questões éticas nessas áreas científicas. As respostas indicam que 70% dos entrevistados relataram não ter usado qualquer documento de ética em IA, enquanto 17% expressaram desconhecimento da existência de documentação para análise de aspectos éticos das ferramentas de IA. Em uma outra análise, foi observado que 83% dos entrevistados admitiram não ter utilizado recursos adicionais para garantir a concepção ou desenvolvimento ético da inteligência artificial (Bano, 2023). Em ambientes de pesquisa, como universidades e laboratórios, são essenciais iniciativas de treinamento e conscientização sobre os impactos éticos do uso das tecnologias de IA em pesquisas científicas.

NÍVEIS DE GOVERNANÇA DE IA

Há uma crescente demanda para implementar regulamentações de IA visando minimizar os riscos para a segurança e preservar os direitos humanos, de forma a garantir que seus benefícios contribuam para a solução de desafios globais, como desenvolvimento econômico, desigualdade e mudança climática. Em resumo, muitos países estão discutindo modelos de

governança de IA que criam estruturas, regras, normas e leis para governar dispositivos, serviços e aplicações de IA. No mundo digital, a governança deriva das ações de diferentes atores que definem políticas e constroem um complexo conjunto de normas formais e informais que organizam as diferentes camadas e processos envolvidos na construção de infraestruturas digitais, plataformas, serviços e aplicações usadas pela sociedade.

Há também uma necessidade premente de desenvolver estruturas de governança para pesquisas científicas que fazem uso de IA, de forma a minimizar consequências que decorrem das atividades de pesquisa que fazem uso das ferramentas de IA (Batool, 2023). A governança da IA busca reconhecer a intrincada interconexão entre ciência, tecnologia e sociedade para compreender e supervisionar as extensas repercussões dos algoritmos de IA sobre indivíduos, comunidades e instituições.

A governança é um conceito com muitos significados, que pode ser implementada através de diferentes arranjos institucionais (Filgueiras, 2022). Um arranjo abrangente para diferentes formas de regulação pode ser estruturado em uma hierarquia de níveis de governança. No nível mais alto da estrutura de governança de IA estão legislações nacionais, que estabelecem o contexto para as regulações de IA em instituições. Por exemplo, o “AI Act” aprovado pela União Europeia define regras para promover uma IA confiável na Europa e garantir que os sistemas de IA respeitem os direitos fundamentais, a segurança e os princípios éticos (European Parliament, 2023). A lei

atribui aplicações de IA a três categorias de risco. Primeiro, as aplicações e sistemas que criam um risco inaceitável. Aplicações de IA para calcular pontuação social a ser usada por governos são proibidas. Segundo, as aplicações de alto risco, como uma ferramenta de análise de currículos que classifica os candidatos a emprego, estão sujeitas a requisitos legais específicos. Por fim, as aplicações não explicitamente proibidas ou listadas como de alto risco são em grande parte deixadas sem regulamentação.

No nível intermediário de regulação de IA estão princípios e boas práticas para o desenvolvimento de sistemas responsáveis de IA. Um bom exemplo de iniciativas de governança nesse nível intermediário é a Ordem Executiva sobre o Desenvolvimento Seguro, Responsável e Confiável da Inteligência Artificial do governo americano, publicada em outubro de 2023 (Biden, 2023), que estabelece princípios e prioridades para orientar o governo sobre o desenvolvimento e usos da IA. A Ordem Executiva atribui a mais alta urgência à governança do desenvolvimento e uso da IA de forma segura e responsável, e promove uma abordagem coordenando diferentes setores do governo para alcançar o objetivo de uma IA global.

Em uma abordagem *bottom-up* para governança, o nível inicial da hierarquia de governança de IA começa com regras, normas e políticas para uso de ferramentas de IA dentro de cada instituição. É importante ressaltar que a governança de IA em uma instituição deve estar alinhada com os princípios, práticas e legislações especificadas nos níveis superiores, ou seja, legislação nacional determinada por

Poder Executivo ou Legislativo. É justamente o nível inicial de regulação que laboratórios, universidades e institutos de pesquisa devem adotar para implantação das tecnologias de IA. Nesse primeiro nível, as organizações de pesquisa têm oportunidade de estabelecer regras e normas, quando ainda não se tem no país legislações aprovadas por Congresso e governo, que usualmente necessitam de mais tempo para formar arranjos políticos requeridos para aprovação de legislações. Esse deve ser o ponto de partida das instituições de pesquisa e ensino superior no Brasil para avançar responsavelmente no uso e aplicação das tecnologias de IA.

CONCLUSÃO

Existem muitas definições de governança digital, sem consenso na literatura acadêmica. Conforme proposto por Filgueiras e Almeida (2020), a governança digital é a capacidade de instituições policêntricas no mundo digital de governar (de maneira legítima, inclusiva e segura) o uso dos bens digitais comuns para produzir serviços sustentáveis e políticas públicas implementadas por governos e empresas de maneira não territorial e orientada por resultados. IA responsável é um indicativo de medidas preventivas que buscam compor uma estrutura de governança com múltiplas instituições, princípios, políticas, ferramentas e processos, de modo a garantir que os sistemas de IA sejam desenvolvidos em conformidade com as leis e princípios éticos fundamentais. A governança da IA responsável deve fortalecer na prática valores essenciais como

transparência, justiça, explicabilidade, privacidade e segurança.

A inteligência artificial responsável ajudará a minimizar consequências indesejáveis e não intencionais na pesquisa científica, como, por exemplo, danos à imagem e credibilidade da ciência. À medida que a IA permeia várias áreas das ciências, gerenciar os riscos se torna cada vez mais importante. Princípios para uma IA responsável são um ponto de partida, mas só fornecerão o que as instituições precisam se forem combinados com práticas de governança que ajudem efetivamente a guiar o desenvolvimento e uso de aplicações de IA.

O avanço da IA nas instituições de ensino superior no Brasil deverá ser acompanhado de um processo de governança, incluindo documentação e ferramentas de apoio aos pesquisadores e professores, de modo a ser eticamente responsável. Operacionalizar diretrizes de IA responsável significa transformar princípios abstratos em ações concretas, processos e práticas que podem ser incorporados aos sistemas nos laboratórios e universidades. Isso aumenta a transparência e a confiança entre os usuários e partes interessadas, diminuindo a distância entre os conceitos teóricos e sua aplicação prática na IA.

REFERÊNCIAS

- BAEZA-YATES, R.; VILLOSLADA, P. *Human vs. artificial intelligence*. IEEE 4th International Conference on Cognitive Machine Intelligence (CogMI). 14/dez./2022, pp. 40-8.
- BANO, M. et al. "Investigating responsible AI for scientific research: an empirical study". arXiv preprint arXiv:2312.09561, 2023.
- BATOOL, A.; ZOWGHI, D.; BANO, M. "Responsible AI governance: a systematic literature review". arXiv preprint arXiv:2401.10896, 2023.
- BIDEN, J. R. *Executive order on the safe, secure, and trustworthy development and use of artificial intelligence*. 2023.
- COUNCIL, US Public Policy. "Statement on principles for responsible algorithmic systems". *Communications of the ACM*, 2022.
- DIGNUM, V. "Responsible autonomy". arXiv preprint arXiv:1706.02513, 2017, pp. 4.698-704.
- EITEL-PORTER, R. "Beyond the promise: implementing ethical AI". *AI and Ethics*, v. 1, 2021, pp. 73-80.

- EUROPEAN PARLIAMENT. *Artificial Intelligence Act: deal on comprehensive rules for trustworthy AI*. 2023. Disponível em: <https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai>. Acesso em: 3/mar./2024.
- FILGUEIRAS, F.; ALMEIDA, V. "Governance for the digital world: neither more state nor more market". *Springer Nature*, 2020.
- KITANO, H. "Nobel Turing Challenge: creating the engine for scientific discovery". *npj Systems Biology and Applications*, v. 7, n. 1, 2021, p. 29.
- MENDONÇA, R. F.; ALMEIDA, V.; FILGUEIRAS, F. *Algorithmic institutionalism: the changing rules of social and political life*. Oxford, Oxford University Press, 2023.
- MOURA, M. "Universidades públicas respondem por mais de 95% da produção científica do Brasil". *Academia Brasileira de Ciências*. Online. 2019. Disponível em: <https://www.abc.org.br/2019/04/15/universidades-publicas-respodem-por-mais-de-95-da-producao-cientifica-do-brasil/>. Acesso em: 3/mar./2024.
- NATIONAL ACADEMIES OF SCIENCES, ENGINEERING, AND MEDICINE et al. *Fostering responsible computing research: Foundations and practices*. 2022.
- RAKOVA, B. et al. "Where responsible AI meets reality: practitioner perspectives on enablers for shifting organizational practices". *Proceedings of the ACM on Human-Computer Interaction*, v. 5, n. CSCW1, 2021, pp. 1-23.
- UNIVERSITY OF CALIFORNIA. *Recommendations to guide the University of California's artificial intelligence strategy*. Presidential Working Group on AI, Final Report, out./2021.
- VAN NOORDEN, R.; PERKEL, J. M. "AI and science: what 1,600 researchers think". *Nature*, v. 621, n. 7980, 2023, pp. 672-5.
- WANG, H. et al. "Scientific discovery in the age of artificial intelligence". *Nature*, v. 620, n. 7972, 2023, pp. 47-60.
- YEUNG, K. "Recommendation of the Council on Artificial Intelligence (OECD)". *International Legal Materials*, v. 59, n. 1, 2020, pp. 27-34.

Inteligência artificial e mudanças climáticas

Paulo Artaxo

Luciana Varanda Rizzo

Luiz Augusto Toledo Machado

resumo

IA é cada vez mais imprescindível para o enfrentamento das mudanças do clima. Imagens de satélite e sensores produzem uma quantidade gigantesca de dados a cada segundo, que são compilados e analisados com ferramentas de IA. Modelos de previsão de tempo e modelos climáticos globais, de parametrização física altamente complexa, são explicitamente calculados com IA, que também otimiza as fontes de energia renováveis e as redes de distribuição de energia elétrica. Apesar dessas aplicações, desafios ligados à ética, à privacidade dos dados e ao impacto ambiental do treinamento da IA exigem atenção cuidadosa. O equilíbrio entre tecnologia e ética é crucial para garantir o uso responsável da IA na luta contra as alterações climáticas.

Palavras-chave: mudanças climáticas; modelos de previsão do tempo; IA e energias renováveis; Amazônia; ética e clima.

abstract

AI is increasingly essential for tackling climate change. Satellite images and sensors produce a huge amount of data every second. They are compiled and analyzed with AI tools. Forecast models, weather, and global climate models, with highly complex physical parameterization, are explicitly calculated with AI, which also optimizes energy sources, renewable energy, and electrical energy distribution networks. Despite these applications, challenges linked to ethics, data privacy, and the environmental impact of AI training require careful attention. The balance between technology and ethics is crucial for ensuring the responsible use of AI in the fight against climate change.

Keywords: *climate change; weather forecast models; AI and renewable energy; Amazon; ethics and climate.*

Dentre os vários desafios que nossa sociedade enfrenta atualmente está a questão das mudanças climáticas (IPCC, 2023). A necessidade urgente de reduzir emissões de gases de efeito estufa, e ao mesmo tempo adaptar nossa sociedade ao clima que já mudou, traz desafios socio-

econômicos e científicos importantes, especialmente para países em desenvolvimento (Artaxo, 2012). Esses desafios são complexos, com impactos em nossa estrutura social, econômica e ambiental.

Neste contexto, a inteligência artificial emergiu como uma ferramenta poderosa no esforço global para enfrentar e mitigar os impactos das alterações climáticas (UNDP, 2023). À medida que o nosso pla-

queta enfrenta desafios ambientais sem precedentes, as tecnologias de IA podem ser utilizadas em vários setores para melhorar a nossa compreensão sobre as alterações climáticas, otimizar a utilização de recursos e desenvolver soluções inovadoras para um futuro sustentável.

Um dos grandes desafios da ciência das mudanças climáticas é a complexidade do sistema terrestre, onde tudo está interligado. O clima do planeta depende das trocas de calor que acontecem entre

PAULO ARTAXO é professor titular do Instituto de Física da USP, coordenador do Centro de Estudos Amazônia Sustentável e vice-presidente da Sociedade Brasileira para o Progresso da Ciência (SBPC).

LUCIANA VARANDA RIZZO é professora do Instituto de Física da USP.

LUIZ AUGUSTO MACHADO é professor visitante do Instituto de Física da USP.

a atmosfera, os oceanos e os continentes. Por exemplo, o aquecimento das águas em uma parte do Oceano Pacífico afeta o padrão de circulação dos ventos, alterando a distribuição de umidade e de calor em toda a zona tropical do planeta, em um fenômeno conhecido como El Niño. Um outro exemplo é o desflorestamento na Amazônia, que tem causado mudanças significativas na temperatura e no regime de chuvas da região, além de afetar o transporte de umidade para as regiões agrícolas do Centro-Oeste do Brasil (Aragão, 2012). Assim, uma mudança local pode ter consequências regionais e até globais. A inteligência artificial pode contribuir para desvelar padrões e interconexões entre processos que ainda não são evidentes devido à complexidade do sistema terrestre.

Com o avanço tecnológico dos últimos 50 anos, a quantidade de dados sobre o sistema terrestre cresceu exponencialmente. Uma gigantesca quantidade de dados é gerada a cada segundo, com medidas de centenas de satélites e redes de observações na superfície, além de produtos de modelagem numérica. É bom ter muitos dados disponíveis, porque isso permite que os cientistas acompanhem as mudanças que têm acontecido no planeta. Porém, essa enorme quantidade de dados não tem significado algum se não for compilada, validada, integrada e reduzida em seu escopo, para ser útil na definição das melhores estratégias no enfrentamento das mudanças climáticas. Uma das principais contribuições da IA reside na sua capacidade de processar grandes quantidades de dados com rapidez. Os cientistas climáticos utilizam métodos e algoritmos de IA para analisar dados climáticos, imagens de satélite e dados de senso-

res, fornecendo informações valiosas sobre padrões e tendências regionais e globais. Os algoritmos de aprendizagem de máquina podem identificar relações complexas em conjuntos de dados climáticos, ajudando a melhorar os modelos e previsões climáticas. Esta melhor compreensão é crucial para que políticas públicas sobre questões climáticas e ambientais sejam elaboradas na melhor ciência disponível, fornecendo soluções para a adaptação climática e estratégias de mitigação.

A IA também é fundamental no monitoramento e gestão sustentável dos recursos naturais. Por exemplo, sistemas alimentados por IA podem analisar dados de satélite para monitorar desmatamento de florestas e mudanças no uso da terra praticamente em tempo real, como, por exemplo, o Dynamic World (Brown et al., 2022; <https://dynamicworld.app>). A IA também é utilizada para avaliar a saúde dos ecossistemas, quantificando riscos associados à dispersão de poluentes, detectando a proliferação de algas em ecossistemas aquáticos, prevenindo a toxicidade de novos compostos ao meio ambiente e dando suporte a projetos de restauração ecológica (Cui et al., 2023). Ao identificar padrões e anomalias, a IA permite esforços de conservação mais eficazes e práticas sustentáveis de gestão de terras.

No domínio das energias renováveis, a IA auxilia na otimização da eficiência e confiabilidade das fontes de energia renováveis. As redes inteligentes de distribuição de energia elétrica, alimentadas por algoritmos de IA, podem gerir a distribuição de energia, prever padrões de demanda e integrar fontes de energia renováveis nas infraestruturas existen-

tes (Thirunavukkarasu et al., 2023). As tecnologias baseadas em IA melhoram o desempenho dos parques solares e eólicos, prevendo padrões climáticos e otimizando a produção de energia em conformidade com a disponibilidade de ventos e radiação solar. Por exemplo, o Centro Nacional de Pesquisas Atmosféricas dos EUA realiza previsões que dão suporte à indústria de energia solar e eólica, contribuindo para mitigar riscos associados a mudanças nas condições atmosféricas em curto e longo prazos (NCAR, <https://ral.ucar.edu/projects/ncars-contribution-to-wind-and-solar-energy-prediction>).

De forma mais geral, a IA pode contribuir para o desenvolvimento de soluções inovadoras para combater as alterações climáticas. Desde a concepção de edifícios mais eficientes em termos energéticos até a criação de tecnologias avançadas de captura de carbono, a IA pode promover a implementação de práticas sustentáveis em muitos processos industriais. A IA pode ser um fator de mudança na agricultura, oferecendo soluções inovadoras para enfrentar os desafios colocados pelas alterações climáticas. À medida que o nosso planeta sofre mudanças nos padrões climáticos, a integração das tecnologias de IA na agricultura torna-se cada vez mais essencial para a produção alimentar sustentável e com baixas emissões. Uma das principais contribuições da IA na agricultura sustentável é na gestão de recursos. O aumento de eventos climáticos extremos está impactando diretamente a produtividade agrícola. Os sistemas alimentados por IA analisam dados climáticos, condições do solo e saúde das culturas em tempo real, fornecendo aos agricultores infor-

mações práticas para tomarem decisões mais bem embasadas. A agricultura de precisão, possibilitada pela IA, permite irrigação direcionada, fertilização precisa e uso otimizado de pesticidas, minimizando o impacto ambiental e garantindo a utilização eficiente dos recursos. Algoritmos de aprendizado de máquina analisam dados meteorológicos históricos, desempenho das colheitas e outros fatores relevantes para prever tendências climáticas futuras. Essa capacidade de previsão auxilia no planejamento da melhor época de plantio e colheita, na administração de rotações de culturas e na mitigação dos riscos associados a eventos climáticos extremos, aumentando, em última análise, a resiliência em face das incertezas climáticas.

A IA contribui para o desenvolvimento de culturas resistentes ao clima através de técnicas avançadas de melhoramento. Ao analisar vastos conjuntos de dados relacionados com a genética e características das plantas, os algoritmos de IA podem identificar marcadores genéticos associados à tolerância a condições climáticas específicas. Essa informação auxilia no desenvolvimento de culturas mais resistentes a seca, pragas e doenças, contribuindo para a criação de ecossistemas agrícolas robustos e sustentáveis. A escassez de água no Brasil Central é uma consequência significativa das mudanças climáticas e é necessário o desenvolvimento de sistemas de irrigação alimentados por IA que ofereçam soluções eficientes de gestão da água. Esses sistemas monitoram os níveis de umidade do solo, as previsões meteorológicas e as necessidades hídricas das culturas para fornecer quantidades precisas de água onde e quando necessário.

Ao minimizar o desperdício de água, a IA ajuda a conservar esse recurso vital e promove práticas agrícolas sustentáveis. No Brasil, estudos da Embrapa (2018) dão suporte a inúmeras iniciativas de aplicação de técnicas de IA na agricultura.

Apesar dessas aplicações promissoras, os desafios permanecem. Considerações éticas, preocupações com a privacidade dos dados e o impacto ambiental do treinamento de grandes modelos de IA são áreas que requerem atenção cuidadosa de nossa sociedade e da ciência. Encontrar um equilíbrio entre os avanços tecnológicos e as práticas éticas é crucial para garantir a implantação responsável da IA na luta contra as mudanças climáticas.

Em conclusão, a integração da inteligência artificial nas iniciativas relativas às alterações climáticas tem o potencial de revolucionar a nossa abordagem aos desafios ambientais. Ao aproveitar o poder analítico da IA, podemos obter conhecimentos mais profundos sobre os padrões climáticos, otimizar a gestão de recursos e desenvolver soluções sustentáveis. No entanto, é imperativo abordar a implantação da IA com um compromisso com considerações éticas e responsabilidade ambiental, garantindo que os benefícios da IA contribuam positivamente para um futuro mais sustentável e resiliente.

CONTRIBUIÇÕES DA IA NO APRIMORAMENTO DOS MODELOS CLIMÁTICOS GLOBAIS

As previsões do tempo e do clima têm sido cruciais na história da humanidade para permitir planejamento agrícola eficaz, salvaguardando contra desastres naturais

e facilitando estratégias e tomada de decisões em diversos setores. Nesse contexto, a necessidade de previsões meteorológicas e climáticas precisas é altamente significativa. Os complexos modelos chamados *Earth System Models* (ESM) integram as interações entre a atmosfera, oceano, ecossistemas terrestres, gelo e biosfera, em que as emissões e absorções naturais e antrópicas são levadas em conta. Em particular, os elos de retroalimentação ou *feedbacks* entre as componentes são extremamente complexos (Kawamiya et al., 2020). O impacto das ações humanas e políticas também tem que ser levado em conta, com toda a sua complexidade.

Os meteorologistas denominam de “revolução silenciosa” uma melhoria gradual, mas constante, na previsão do tempo. Hoje, a previsão de tempo para seis dias é quase tão boa quanto a previsão de três dias de 30 anos atrás. Menos frequentemente do que no passado, tempestades severas ou ondas de calor pegam as pessoas desprevenidas. Essa revolução salvou vidas e dinheiro, mas também tem um custo: milhões de dólares em supercomputadores produzindo essas previsões.

A inteligência artificial está estimulando outra revolução na previsão numérica do tempo (Voosen, 2023). Em poucos minutos, em computadores desktop, sistemas de IA treinados podem agora fazer previsões similares aos melhores modelos tradicionais. A principal agência meteorológica do mundo, o Centro Europeu de Previsões Meteorológicas de Médio Prazo (ECMWF), adotou a tecnologia: recentemente começou a gerar as suas próprias previsões experimentais de IA. Os algoritmos estão permitindo previsões mais

frequentes a liberação de caros recursos computacionais para outros problemas mais espinhosos (Chen et al., 2023).

Os gigantes da tecnologia também estão muito ativos nesta área. O sistema Google DeepMind, usando o modelo GraphCast, e a Huawei, com o modelo Pangu-Weather, são importantes novidades nesta área. O Google tem um modelo meteorológico de IA de curto prazo que faz previsões contínuas de 24 horas tão precisas quanto as de qualquer agência meteorológica. Os modelos meteorológicos tradicionais começam por alimentar um instantâneo das condições atuais, com base em observações de satélites, estações meteorológicas, balões e boias oceânicas, em complexos modelos computacionais que dividem a superfície do planeta e a atmosfera em milhões de grades, com tamanhos de 1 a 10 quilômetros. Em cada ponto dessa grade são resolvidas as leis físicas da dinâmica dos fluidos, termodinâmicas e leis de conservação, com grande custo computacional. Os modelos climáticos podem levar várias horas para serem executados em supercomputadores com milhões de processadores.

Os novos modelos utilizando IA evitam esses complexos cálculos matemáticos em favor do chamado “aprendizado profundo”, ou *deep learning* (Reichstein et al., 2019). Eles identificam padrões de como a atmosfera evolui naturalmente, após extenso treinamento utilizando mais de 40 anos de dados de “reanálise” do ECMWF, que é uma combinação de observações e previsões de modelos de curto prazo que representa a melhor e mais completa imagem do clima passado.

Quando inicializado com um estado inicial da atmosfera com base na mesma

combinação de observações e modelagem, o sistema GraphCast (Lam et al., 2024) pode superar a previsão do ECMWF em até dez dias em 90% de suas metas de verificação, incluindo trajetórias de furacões e temperaturas extremas. O processo de aprendizado de máquina foi extenso, utilizou 32 supercomputadores e quatro semanas para construir o modelo de IA e 36,7 milhões de parâmetros para ser construído. No modelo de IA, o algoritmo final resultante é leve o suficiente para funcionar em menos de um minuto em um único computador desktop. Resultados similares foram observados com o sistema Pangu-Weather, alimentando-o apenas com as observações que fazem parte do modelo meteorológico operacional do ECMWF.

No Brasil, pesquisadores têm aplicado técnicas de IA para realizar previsões de tempo em curto prazo (*nowcasting*). Pesquisadores do Laboratório Associado de Computação e Matemática Aplicada (LAC) do Instituto Nacional de Pesquisas Espaciais (Inpe) desenvolvem ferramentas computacionais que permitem prever a ocorrência e a intensidade de tempestades e de relâmpagos com até três horas de antecedência, aplicando técnicas de IA a imagens de radares meteorológicos. Também existem pesquisas no sentido de monitorar riscos de enchentes e alertar os moradores, como o sistema E-Noé, desenvolvido por pesquisadores do Instituto de Ciências Matemáticas e de Computação (ICMC) da Universidade de São Paulo.

Em breve será possível realizar as chamadas previsões de conjuntos (*ensembles*), uma inovação em previsão meteorológica que ajuda a capturar a incerteza ao executar um modelo várias vezes para criar

uma série de resultados possíveis. Os pesquisadores de IA poderiam seguir a técnica tradicional de ajustar as condições climáticas iniciais com pequenas “perturbações”, ou adaptar as técnicas generativas de IA gerando condições ajustadas dinamicamente. Essas previsões conjuntas poderiam ajudar os modelos de IA a prever melhor eventos extremos, como furacões, cuja previsão de intensidade e trajetória atualmente é um desafio. O GraphCast pode abrir novas direções para outros problemas importantes de previsão geoespaço-temporal, incluindo clima e ecologia, energia, agricultura e atividade humana e biológica, bem como outros sistemas dinâmicos complexos. Simuladores treinados em dados ricos e do mundo real serão cruciais para o avanço do papel do aprendizado de máquina nas ciências físicas.

O potencial desses modelos não se limita à previsão do tempo. Eles não podem prever climas futuros, porque os conjuntos de dados de aprendizado ao longo dos últimos 40 anos não são suficientemente longos para captar tendências de aquecimento global, que estão sujeitas a *feedbacks* complexos de nuvens, gases e aerossóis que podem acelerar ou atenuar as alterações climáticas. Mas poderiam ajudar uma nova geração de modelos climáticos de alta resolução que estão sendo desenvolvidos para funcionar em supercomputadores com nível de processamento das Exascale, as mais recentes máquinas ultrarrápidas. Assim que esses modelos produzirem resultados suficientes para treinamento, a IA poderá realizar previsões climáticas para as próximas décadas ou séculos, de modo muito mais fácil e com precisão similar ou melhor.

IA EM PESQUISAS AMAZÔNICAS

Por ser a maior floresta tropical do mundo, a Amazônia desempenha um papel importante no equilíbrio climático do planeta. A floresta armazena e remove carbono da atmosfera através da fotossíntese e transpira vapor d’água, influenciando o regime de chuvas. Dessa forma, o clima e a vegetação estão intimamente relacionados (Artaxo et al., 2023). O desflorestamento e as mudanças climáticas diminuem a capacidade da floresta de remover carbono da atmosfera, além de torná-la mais suscetível à degradação. Estudos apontam a existência de um ponto de não retorno, em que a capacidade de regeneração da floresta seria insuficiente para resistir às pressões do desflorestamento e das mudanças climáticas.

Para prever o que pode acontecer com a Amazônia e com o clima no futuro, os cientistas desenvolvem modelos matemáticos que simulam as interações entre a floresta e a atmosfera com base no conhecimento atual. Por exemplo, sabe-se que em dias ensolarados ocorre mais fotossíntese, desde que a vegetação tenha pleno acesso à água e a nutrientes. Esse conhecimento é traduzido em equações matemáticas que constituem um modelo. Apesar dos avanços na ciência, nosso conhecimento ainda é limitado, isto é, não sabemos como toda a variedade de espécies vegetais da Amazônia responde a mudanças de temperatura, nebulosidade, disponibilidade de água, concentração de CO₂ atmosférico e brotamento de novas folhas. Em outras situações, podemos até ter conhecimento sobre os processos físicos, químicos e biológicos, mas representá-los detalhadamente em modelos matemáticos resultaria em um alto custo computacional.

É nesse ponto que a IA, utilizada em conjunto com modelos matemáticos tradicionais, pode aprimorar as previsões. Técnicas de aprendizagem de máquina podem ser utilizadas para reconhecer padrões, avaliar a sensibilidade de diferentes preditores e identificar relações de causa e consequência ainda desconhecidas. Para isso, é necessário um extenso conjunto de dados, contendo um grande volume de perguntas (preditores) e de respostas corretas (variável-alvo), para que o algoritmo possa aprender.

A IA também pode contribuir para aumentar a disponibilidade de dados sobre a Amazônia. O bioma se estende por 4 milhões de km², sendo que as áreas mais preservadas são de difícil acesso. É um grande desafio manter programas de monitoramento de taxas de fotossíntese, estoques de biomassa e degradação florestal em uma área tão extensa. Algoritmos de IA aplicados a dados de satélite e a dados de modelos climáticos globais têm contribuído para preencher lacunas de observações em áreas de difícil acesso na Amazônia (Jung et al., 2020). Imagens de diferentes satélites têm sido combinadas em modelos de aprendizagem de máquina para estimar parâmetros que indicam o estado de saúde da floresta, como índices de área foliar, estoque de biomassa, entre outros (Estrada et al., 2023).

Outro desafio importante é monitorar a degradação florestal causada pela remoção de árvores de madeira nobre, fogo sob o dossel, efeitos de borda e pequenas estradas de terra. A degradação afeta a capacidade da floresta de remover carbono da atmosfera, além de diminuir a sua resiliência às mudanças climáticas e aos eventos climáticos extremos. É um desafio monitorar a degradação florestal via satélite, pois muitas

vezes o dano ocorre abaixo do dossel da floresta, sendo imperceptível para os sensores. Outras vezes, os sensores via satélite não possuem resolução suficiente para detectar clareiras de poucos metros quadrados de área. Modelos de aprendizagem profunda têm sido bem-sucedidos na identificação de diferentes classes de degradação florestal a partir de dados de satélite em alta resolução espacial (Dalagnol et al., 2023).

Métodos de IA e de visão computacional também têm sido aplicados em dados de uso e cobertura da terra para reconhecer padrões de desmatamento e identificar suas principais causas. Redes neurais convolucionais são utilizadas para prever padrões de desmatamento na Amazônia, produzindo mapas que indicam o risco de que um determinado pixel de floresta evolua para um pixel de desmatamento (Ball et al., 2022). Caracterizar os processos que levam ao desflorestamento e identificar áreas de maior risco são fundamentais para desenvolver, implementar e aprimorar políticas de conservação e manejo de áreas florestais. Dessa forma, o uso de IA pode contribuir para a preservação da floresta amazônica. Outra aplicação interessante é o uso de IA para avaliar a recuperação da biodiversidade em áreas de restauração florestal, com base em informações bioacústicas (Müller et al., 2023).

CONCLUSÕES E PERSPECTIVAS FUTURAS

O uso de técnicas de IA nas ciências do sistema terrestre é muito promissor, tendo em vista o grande volume e variedade de dados disponíveis para promover

o aprendizado de modelos estatísticos. Em algumas situações, técnicas de aprendizagem de máquina podem representar de maneira eficiente relações complexas entre as diversas variáveis do sistema terrestre, com um custo computacional menor em comparação aos modelos matemáticos tradicionais. Técnicas de IA também são eficientes para reconhecer padrões em imagens de satélite, e muitas vezes podem desvelar novas interações entre variáveis, contribuindo para o avanço do conhecimento.

Entretanto, é importante considerar as limitações das técnicas de IA. Em primeiro lugar, o aprendizado de máquina é construído com base em grandes volumes de dados, que geralmente foram coletados no passado. Assim, os modelos de IA podem reproduzir bem condições já observadas anteriormente, mas podem falhar na representação de situações atípicas ou sob condições de mudança. Nessas situações, os modelos matemáticos tradicionais, baseados em conhecimento físico, podem ter um desempenho superior aos modelos de aprendizagem de máquina. Outro ponto fraco de algoritmos de aprendizagem de máquina convencionais é a dificuldade de extrair conhecimento dos modelos. Muitas vezes, os modelos de IA funcionam como “caixas-pretas”, capazes de fazer previsões acuradas, mas que não explicam por que uma determinada combinação de variáveis contribui para um determinado desfecho. Novas técnicas de IA têm sido desenvolvidas para facilitar a extração de conhecimento e aumentar a precisão e previsibilidade dos modelos estatísticos.

Um artigo recente (Carlsen et al., 2024) demonstra que a inteligência arti-

ficial alcançou um ponto de inflexão, com implicações significativas para a forma como a pesquisa precisa abordar as tecnologias emergentes e como elas impulsionam o desenvolvimento socioeconômico de longo prazo e os cenários de mudanças climáticas. A inteligência artificial é sem dúvida um fator que irá modificar a nossa sociedade, uma vez que ela vai apresentar soluções sem a necessidade de realizar toda uma cadeia de raciocínios sobre o assunto. A IA pode trazer grandes benefícios, que vão desde a democratização de alguns complexos processos como, por exemplo, fazer uma previsão de tempo para os próximos três dias (que hoje normalmente demanda um centro de supercomputação, que corresponde a um investimento de pelo menos US\$ 50 milhões), que poderá ser feita em seu laptop, com igual ou melhor qualidade, utilizando ferramentas como o GraphCast (Lam et al., 2024).

Um ponto a refletir é a capacidade de os modelos de IA conseguirem reproduzir algo que não existe no histórico de dados, algo novo, como os efeitos não lineares das mudanças climáticas, ou o surgimento de novos processos físicos nos quais o sistema não realizou o aprendizado. Contudo, essa limitação não é fundamental considerando-se o enorme número de processos que ainda não conseguimos modelizar, ou seja, entender, permitindo prevê-los mesmo sem o conhecimento científico da complexa interrelação de processos em diferentes escalas. Ao mesmo tempo, preocupa o fato de se aceitar o resultado sem entender por que ele ocorre. É possível que a IA afaste o interesse no entendimento dos processos

em detrimento de respostas rápidas e sem esforços. Qual será a consequência disso? Perderemos o poder de transformar os enigmas da natureza em equações diferenciais e passaremos a aceitar os resultados puramente estatísticos, como hoje é tratada a física quântica?

A IA leva a uma série de questões que podem ser debatidas calorosamente, mas

uma área em que ela atua pode trazer enormes consequências às sociedades – a área que envolve o papel do homem na natureza. A área de ciências sociais é muito mais difícil do que determinar os resultados de processos físicos, químicos ou biológicos. Temos muito o que discutir para normatizar o uso de IA para o bem da sociedade, tarefa mais que necessária.

REFERÊNCIAS

- ARTAXO, P. "Break down boundaries in climate research". *Nature, World View Section*, v. 481, 2012, p. 239. Disponível em: <http://www.nature.com/news/break-down-boundaries-in-climate-research-1.9844>.
- ARTAXO, P. "Amazon deforestation implications in local/regional climate change". *PNAS - Proceedings of the National Academy of Sciences*, v. 120, n. 50, e2317456120, 2023. Disponível em: <https://doi.org/10.1073/pnas.2317456120>.
- ARAGÃO, L. E. O. C. et al. "The rainforest water pump". *Nature*, v. 489, 2012, pp. 217-8. Disponível em: <https://www.nature.com/articles/nature11485>.
- BALL, J. G. et al. "Using deep convolutional neural networks to forecast spatial patterns of Amazonian deforestation". *Methods in ecology and evolution*, v. 13, n. 11, 2022, pp. 2.622-34.
- BROWN, C. F. et al. "Dynamic world, near real-time global 10m land use land cover mapping". *Sci Data*, n. 9, 2022, p. 251. Disponível em: <https://doi.org/10.1038/s41597-022-01307-4>.
- CARLSEN, H. et al. "Chasing artificial intelligence in shared socioeconomic pathways". *One Earth*, v. 7, 2024, pp. 18-22. Disponível em: <https://doi.org/10.1016/j.oneear.2023.12.015>.
- CHEN, L. et al. "Machine learning methods in climate prediction: a survey". Disponível em: <https://www.preprints.org/manuscript/202309.1764/v1>.

- CUI, S. et al. "Advances and applications of machine learning and deep learning in environmental ecology and health". *Environmental Pollution*, 2023.
- DALAGNOL, R. et al. "Mapping tropical forest degradation with deep learning and Planet NICFI data". *Remote Sensing of Environment*, v. 298, 2023.
- EMBRAPA. *Visão 2030: o futuro da agricultura brasileira*. Brasília, 2018. Disponível em: <https://www.embrapa.br/documents/10180/9543845/Vis%C3%A3o+2030+-+o+futuro+da+agricultura+brasileira/2a9a0f27-0ead-991a-8cbf-af8e89d62829>.
- ESTRADA, J. S. et al. "Machine learning assisted remote forestry health assessment: a comprehensive state of the art review". *Frontiers in Plant Science*, v. 14, 2023.
- IPCC 2023. *Climate change 2023: Synthesis report*. Contribution of working groups I, II, and III to the sixth assessment report of the Intergovernmental Panel on Climate. Genebra, 2023, pp. 35-115. Disponível em: <https://www.ipcc.ch/report/ar6/syr/>.
- JUNG, M. et al. "Scaling carbon fluxes from eddy covariance sites to globe: synthesis and evaluation of the Fluxcom approach". *Biogeosciences*, v. 17, 2020, pp. 1.343-65. Disponível em: <https://doi.org/10.5194/bg-17-1343-2020>.
- KAWAMIYA, M. et al. "Two decades of Earth system modeling with an emphasis on the Model for Interdisciplinary Research on Climate (Miroc)". *Prog Earth Planet Sci*, v. 7, n. 64, 2020. Disponível em: <https://doi.org/10.1186/s40645-020-00369-5>.
- LAM et al. "Learning skillful medium-range global weather forecasting". *Science*, v. 382, 2023, pp. 1.416-21.
- MÜLLER, J. et al. "Soundscapes and deep learning enable tracking biodiversity recovery in tropical forests". *Nature Communications*, v. 14, n. 1, 2023, p. 6.191.
- REICHSTEIN, M. et al. "Deep learning and process understanding for data-driven Earth system science". *Nature*, v. 566, n. 7.743, 2019, pp. 195-204.
- THIRUNAVUKKARASU, M.; SAWLE, Y.; LALA, H. "A comprehensive review on optimization of hybrid renewable energy systems using various optimization techniques". *Renewable and Sustainable Energy Reviews*, v. 176, 2023.
- UNDP 2023. Disponível em: <https://unfccc.int/news/ai-for-climate-action-technology-mechanism-supports-transformational-climate-solutions>.
- VOOSEN, P. "AI is set to revolutionize weather forecasts". *Science in Depth*, 17/nov./2023.

Inteligência artificial em saúde

Edson Amaro Jr.

Helder Nakaya

Luiz Vicente Rizzo

resumo

IA em saúde é uma realidade. Suas possíveis aplicações nas áreas de diagnóstico, pesquisa e interface entre provedores e pacientes são amplas e algumas já estão em uso, mesmo sem a certeza de que, em sua configuração atual, apresentem vantagens para os pacientes, provedores ou para a sociedade. As possibilidades, principalmente para IA generativa, são imensas. Entretanto, estudos apropriados devem ser conduzidos para validar o seu uso, não diferentes daqueles que se aplicam a qualquer avanço na área de saúde. Os próximos meses e anos trarão avanços no uso e nas fronteiras dessa tecnologia. Neste texto descrevemos algumas das utilidades e possibilidades da IA em saúde, seja na sua entrega, em pesquisa ou na gestão.

Palavras-chave: IA em saúde; IA na pesquisa diagnóstica; medicina ética; IA e relação paciente-saúde.

abstract

AI in healthcare is a reality. Its possible applications in areas such as diagnosis, research and interface between providers and patients are already in use, even without the certainty that, in their current configuration, they present advantages for patients, providers or society. The possibilities, especially for generative AI, are immense. However, rigorous studies must be conducted to validate its use, no different from those that apply to any advancement in healthcare. The near future will bring advances in the use and frontiers of this technology. In this text we describe some of the uses and possibilities of AI in healthcare, whether in its results, in research or in management.

Keywords: *AI in healthcare; AI in diagnostic research; ethical medicine; AI and patient-health relationship.*

A

medicina moderna depende de dados. A decisão por um tratamento ou a confirmação de um diagnóstico são geralmente apoiadas por dados na forma de resultados de exames clínicos, laboratoriais, imagens e até pelo genoma do paciente. Porém, a saúde não pode ser definida apenas pelos dados medidos nesses exames. Para tomar decisões, o médico precisa utilizar não apenas sua experiência, mas também integrar informações que não foram capturadas de forma objetiva em exames. Uma inteligência artificial que toma decisões baseada apenas em alterações de padrões de dados será limitada a doenças que já não necessitam tanto de “inteligência humana”.

Além disso, há várias definições para inteligência – a grande maioria, e por razões

óbvias, trata a inteligência como uma característica antropológica. O cérebro usa roteiros de localização espacial e temporal para a tomada de decisões e se adapta a recompensas por decisões bem-sucedidas. Esses roteiros não se aplicam para a inteligência artificial, pelo menos na forma atual. Isto posto, o uso dessas tecnologias na área de saúde não é novidade e não é reversível. Sua importância é de tal monta que o *New*

EDSON AMARO JR. é médico neurorradiologista do Hospital Albert Einstein e professor associado da Faculdade de Medicina da USP.

HELDER NAKAYA é pesquisador sênior do Hospital Albert Einstein e professor adjunto da Universidade Emory (Atlanta, EUA) e da Faculdade de Ciências Farmacêuticas da USP.

LUIZ VICENTE RIZZO é diretor de Pesquisa do Hospital Albert Einstein e foi professor titular do Instituto de Ciências Biomédicas da USP (2005-2009).

England Journal of Medicine, um dos maiores jornais científicos na área médica, dedica uma revista especial para essas ferramentas¹.

Antes de discorrer sobre os possíveis usos desta tecnologia, é importante pontuar que a maior parte das funcionalidades apresentadas, principalmente na mídia leiga, não foram submetidas ao mesmo escrutínio científico de grande parte dos procedimentos que a medicina moderna utiliza. Essas ferramentas precisam passar por avaliação científica rigorosa, por ensaios clínicos robustos com desenhos apropriados e validados por análise estatística adequada.

Também é importante ressaltar que os modelos de IA dependem diretamente da quantidade e qualidade dos dados, assim como de definições claras dos desfechos que se quer predizer. Se treinarmos modelos com dados insuficientes ou de baixa qualidade (artefatos técnicos, enviesados para certas raças, grupos desbalanceados, confundidores e covariáveis que não são levados em conta etc.), corremos o risco de criarmos uma “burrice artificial” que poderá abalar a confiança nessa poderosa tecnologia, atrasando assim seu desenvolvimento. A definição do desfecho é outra parte importante na criação de um bom modelo de IA. Um desfecho claro é a “morte” (ou “sobrevivência”) de um paciente após uma conduta médica. Porém, muitas vezes o desfecho pode ser relacionado a efeitos adversos a um tratamento com um fármaco, à qualidade de vida do paciente após uma intervenção ou mesmo à gravidade de uma doença. Nesses casos, os critérios utilizados na classificação podem não ser um consenso na comunidade médica e irão

impactar diretamente o poder preditivo dos modelos de IA.

A influência da IA pode ser ampla e será abordada em partes aqui. A saber, na pesquisa, na medicina diagnóstica, na medicina clínica e cirúrgica, na interação profissional de saúde-paciente e na navegação da jornada do paciente.

NA PESQUISA

Um dos maiores impactos da IA será no encurtamento do tempo de desenvolvimento e validação para medicamentos e procedimentos. Desde a descoberta/revisão de epítomos ótimos para a construção de vacinas com a análise tridimensional de proteínas ligantes aos receptores do sistema imune até a previsão de antigenicidade e imunogenicidade das vacinas. E também diminuindo a necessidade do uso de animais, o que em si só é um valor agregado enorme pelo fator moral, mas também pelo econômico, uma vez que a pesquisa pré-clínica hoje carrega uma porcentagem razoável do custo agregado de trazer uma nova medicação ou vacina ao mercado, passando pela triagem de novas moléculas e reposicionamento de medicamentos já desenvolvidos. Um exemplo de sucesso em descoberta de novas drogas usando IA é a empresa Benevolent².

A IA tem o potencial de acelerar a descoberta de medicamentos, permitindo aos pesquisadores analisar rapidamente conjuntos de dados em grande escala, conceber novas moléculas e prever a eficácia de potenciais candidatos a medicamentos em bibliotecas

1 Ver: <https://ai.nejm.org>.

2 Ver: <https://www.benevolent.com/>.

já existentes, ou ainda indicar quais configurações químicas apresentam o maior potencial de evolução para medicamento. Por exemplo, as tecnologias físico-químicas disponíveis que fazem o processo de varredura de elementos da flora e fauna para uso farmacológico são custosas tanto em tempo como em dinheiro. Tendo como ponto de partida elementos da biodiversidade, a IA pode transformar um processo tedioso, caro e de baixo rendimento. Outra área de possível impacto da IA é no desenho de ensaios clínicos e em sua análise. A IA tem sido usada já há algum tempo em pesquisa para analisar grandes conjuntos de dados e identificar padrões que seriam difíceis de serem detectados por seres humanos. Isso permite, por exemplo, que dados de genômica de um indivíduo possam ser analisados rapidamente para a descoberta de suscetibilidades a doenças. Também permite que esses dados sejam usados para melhor determinar se uma coorte em qualquer estudo clínico é representativa do ponto de vista genético da população que se pretende estudar. Definições mais precisas de grupos em um ensaio clínico têm o potencial de diminuir o número de participantes necessários para se obter dados significantes de segurança e eficácia. Também a análise de dados secundários e acessórios por IA tem o potencial de revelar tendências discretas e grupos específicos com efeitos inesperados, ambos importantes achados em qualquer estudo clínico.

Profissionais de saúde, cientistas experimentais e especialistas em IA irão cada vez mais trabalhar juntos em pesquisas acadêmicas relacionados à saúde. Não é à toa que a prestigiada Universidade Johns Hopkins anunciou em 2023 a criação de um novo departamento em IA para saúde

com 100 professores³. Iniciativas similares também acontecem em outras universidades, como Stanford⁴, Mayo Clinic⁵ e University of Oxford⁶. Muitos já entenderam que para se obter grandes avanços científicos será preciso um time verdadeiramente multidisciplinar.

NA MEDICINA DIAGNÓSTICA

Medicina diagnóstica é a área médica/biomédica de avaliação da saúde de um indivíduo por meio de exames laboratoriais, funcionais ou de imagem. O objetivo da medicina diagnóstica é ajudar a determinar a causa de um problema de saúde e fazer um diagnóstico preciso para fornecer um tratamento eficaz. Isso pode envolver várias modalidades, como exames laboratoriais, de imagem (por exemplo, ressonância magnética, tomografia computadorizada etc.), exames anatomopatológicos (biópsias, exames de aspirados e de líquidos corporais), exames funcionais (eletrocardiograma, eletromiografia, prova de função pulmonar etc.). Conquanto um grau de IA já venha sendo usado, principalmente em exames de imagem, a diversidade dos dados de um paciente (dados clínicos, demográficos, de imagem, laboratoriais, ambientais e genéticos, para citar alguns) é uma base intrigante que pode ser analisada. A IA pode analisar grandes quantidades de dados de pacientes, isoladamente qualquer um dos tipos de dados acima, e ainda todos eles combinados, e associá-los

3 Ver: <https://ai.jhu.edu/>.

4 Ver: <https://aihealth.stanford.edu/>.

5 Ver: <https://www.mayo.edu/research/departments-divisions/artificial-intelligence-informatics/overview>.

6 Ver: <https://aideproject.web.ox.ac.uk/>.

e/ou analisá-los em conjunto com informações demográficas, geográficas, de histórico médico e de trabalho. Teoricamente, seja pela maior precisão ou rapidez na análise de um conjunto de dados ou pela possibilidade de análise de todos os dados de uma só vez, a IA poderia fornecer previsões mais precisas de diagnóstico, prognóstico e até tratamento, ajudando os profissionais de saúde a tomar decisões mais embasadas.

Recentemente, a revista *Nature Medicine* publicou uma lista com 11 ensaios clínicos que, segundo os editores, terão em seus dados algumas das descobertas mais importantes do próximo ano⁷. Conquanto esse tipo de previsão – que é mais comum em programas de variedades no domingo à noite – seja de assertividade duvidosa, cumpre lembrar que são ensaios importantes. Em um deles, sobre uso de IA para o diagnóstico precoce de câncer de pulmão, há dois dos 11 ensaios citados em câncer de pulmão. O ensaio randomizado liderado pelo dr. David Baldwin, do Nottingham University Hospitals, segue 150 mil pacientes em seis hospitais do Reino Unido testando a IA aplicada ao raio-X (RX) de pulmão assim que ele é obtido em primeira consulta. A proposta é que essa abordagem possa diminuir o tempo para a indicação da tomografia computadorizada (TC) de tórax e, conseqüentemente, o tempo para o diagnóstico do câncer. Em tempo, o diagnóstico precoce de câncer de pulmão salva muitas vidas. A melhor prática indica que o primeiro teste a ser obtido é

o RX de tórax, seguido pela TC se houver suspeita de câncer. O investigador principal aponta que “a IA está sendo implantada em hospitais sem uma análise adequada dos impactos clínicos, algo que nos preocupa. Se nosso estudo encontrar uma melhora significativa no tempo até o diagnóstico, provavelmente levará a uma mudança imediata no padrão de atendimento para incluir IA no momento da radiografia de tórax”.

Em um futuro próximo, a IA também irá ajudar o médico na hipótese diagnóstica. Além de dados de resultados de exames médicos, os modelos de aprendizado de máquina também serão treinados com dados de textos de prontuário, de *wearables* (anéis e relógios que medem diversos parâmetros de saúde) e de tecnologias genômicas. Grandes modelos de linguagem (do inglês *large language models*), como o que está por trás do ChatGPT, poderão ser usados para interpretar, classificar e organizar o texto livre digitado por médicos em anamneses. As grandes empresas de tecnologia como Google, Samsung, Apple e Meta já começaram a investir em soluções de IA para a saúde. Como detentor de um volume imenso de dados, o SUS se torna uma fonte de ouro para essas empresas. Se investimentos estratégicos não forem feitos em infraestrutura computacional e treinamento de profissionais de IA, poderemos sofrer um colonialismo de dados, no qual venderemos barato nossos dados coletados a duras penas para pagar caro pelo modelo treinado por essas *big techs*.

NA MEDICINA CLÍNICA E CIRÚRGICA

Os sistemas de apoio à decisão clínica alimentados por IA podem fornecer assis-

7 Ver: https://www.nature.com/articles/s41591-023-02699-5?utm_source=linkedin&utm_medium=social&utm_content=organic&utm_campaign=CONR_JRNLS_AWA1_GL_SCON_SME_DA_NATUREPORTFOLIO.

tência e apoio em tempo real para tomar decisões mais informadas sobre o atendimento ao paciente. Alguns já estão em uso há anos por profissionais isolados ou por instituições. No entanto, o desenvolvimento e a implantação da IA no diagnóstico médico ainda estão em fase inicial. Há vários desafios técnicos, éticos e de regularização que devem ser superados para que essas ferramentas possam ser usadas com segurança. Como denotado acima, o escrutínio dessas ferramentas pelo método científico não pode ser relevado apenas porque elas parecem funcionar. O primeiro desafio deve-se à qualidade e disponibilidade dos dados médicos, pois os algoritmos de IA requerem grandes quantidades de dados classificados e de alta qualidade para serem eficazes. Na área de saúde os dados são frequentemente fragmentados, incompletos e impropriamente classificados. Se os dados não forem representativos de um indivíduo ou população, o uso da ferramenta de IA levará a diagnósticos incorretos ou incompletos e ocasionalmente perigosos.

Ainda no grupo dos 11 ensaios clínicos citados acima, há um estudo de aprendizado de máquinas liderado pelo dr. Steven Meex, do Laboratório Diagnóstico Central do Centro Médico de Maastricht, nos Países Baixos. Segundo o pesquisador, “MARS-ED é um ensaio clínico piloto prospectivo, multicêntrico, randomizado, aberto e de não inferioridade de assistência de pontuação de risco para médicos no departamento de emergência. O objetivo é determinar a precisão do diagnóstico, as mudanças nas políticas e o impacto clínico do Risk Index como base para a realização de um ensaio clínico randomizado, multicêntrico e em grande escala”. Há muito entusiasmo sobre

IA. A validação prospectiva de modelos de IA raramente é feita, mas é indispensável, porque a implementação pode ser decepcionante. Se o modelo de IA for validado, os seus efeitos poderão incluir a identificação de pacientes que irão piorar apesar do julgamento clínico. É difícil prever como os médicos reagirão ao serem questionados por um modelo de IA, mas vimos que, embora não mudem frequentemente de ideia, podem tornar-se mais cuidadosos.

O estudo acima se baseia em uma análise retrospectiva realizada no Centro Médico de Maastricht, onde foi introduzido um novo escore de risco clínico baseado em IA, o Risk Index, para prever a mortalidade em 31 dias de pacientes que procuraram tratamento em um pronto-socorro. A ferramenta foi desenvolvida e avaliada em quatro hospitais holandeses, com base em dados de 266.327 pacientes, com 7,1 milhões de resultados laboratoriais disponíveis. O Risk Index superou significativamente em suas previsões a avaliação dos médicos de emergência nas quatro instituições com as variações esperadas quando controlado para anos de experiência do médico, mas sem controles específicos para o desempenho medido desses médicos.

A popularização de sistemas robóticos para cirurgia, principalmente pela diminuição de seus custos, traz consigo a inserção da IA na área cirúrgica. Os consoles atuais de cirurgia robótica já possuem algoritmos para correção de tremor e sistemas de alerta para sangramento, entre outras ferramentas de segurança. Esses sistemas permitem, por exemplo, estender a vida útil de um cirurgião. A integração com programas para a análise de imagens médicas (ressonância magnética, tomografia com-

putadorizada, hemodinâmica etc.) poderá oferecer instrumentos de navegação para o cirurgião e até realizar procedimentos em “piloto automático”.

NA INTERAÇÃO PROFISSIONAL DE SAÚDE-PACIENTE

A interação entre o profissional de saúde e o paciente é muito mais ampla que apenas aquelas ações caracterizadas no momento do atendimento. As ferramentas de IA explicável podem automatizar tarefas rotineiras, liberando os profissionais de saúde para se concentrarem em cuidados mais complexos aos pacientes. Ferramentas de IA que favoreçam a coleta de informações antes, durante e depois da consulta/atendimento podem garantir acurácia, além da otimização do tempo. Ferramentas que ajudem na leitura e composição de comunicações por escrito também têm um papel importante na preservação do tempo útil de pacientes e profissionais de saúde. Aqui há uma nota de cautela: essas ferramentas diminuem a pessoalidade da comunicação. Algumas já estão presentes em plataformas comerciais para qualquer tipo de negócio.

Portanto, a otimização do tempo do médico deve ser investida na qualidade da interação humana entre ele e o paciente. Não deve servir para adicionar mais pacientes no mesmo tempo de consulta, mas permitir um atendimento mais longo, empático e atencioso. Já existem, por exemplo, iniciativas que visam “escutar” e transcrever a consulta em tempo real, colocando as informações diretamente no prontuário médico. Isso livraria o médico de digitar enquanto atende um paciente. Obviamente, a tecnologia avança

muito mais rápido do que nossa capacidade de discutir seu uso ético e responsável.

NA NAVEGAÇÃO DA JORNADA DO PACIENTE

A jornada de um paciente pelos diversos serviços de saúde, especialmente para aqueles com doenças crônicas, pode ser complexa e entediante. Nela, a marcação de consultas, exames, execução de pagamentos, aquisição de medicamentos e outros insumos já podem ser executados por diversas formas de IA em diferentes tipos de equipamento. A dificuldade de comunicação entre as diversas plataformas é um problema ainda pela sua diversidade. Além disso, o acesso rápido à satisfação do paciente e familiares, gerando ajuste da jornada para melhorar sua experiência no atendimento em curso, é outra área em que a IA pode ter um grande impacto.

Ainda há à nossa frente mudanças mais radicais. Num futuro não tão distante, num *shopping center*, supermercado ou farmácia próximos, a saúde poderá estar dentro de uma caixa⁸. O anúncio de 100 milhões de dólares investidos nessa possibilidade mostra o que se está pensando. Tecnologias de IA mais avançadas estão cada vez mais sendo colocadas na pesquisa, como a IA quântica, para acelerar o processo de treinamento convencional e fornecer modelos de diagnóstico rápido. Considerando que os computadores quânticos têm significativamente mais poder de processamento do que os computadores

8 Ver: <https://www.forbes.com/sites/katiejennings/2023/11/15/an-ai-doctor-in-a-box-coming-to-a-mall-near-you/?sh=309234f24284>.

clássicos, isso pode permitir que algoritmos de IA quântica analisem grandes quantidades de dados de saúde em tempo real, talvez levando a diagnósticos mais precisos e de forma mais eficiente. Também a escolha do melhor curso de tratamento para um paciente com base em seu histórico médico e outros fatores pode ser impactada positivamente. Ao utilizar algoritmos de IA para analisar grandes quantidades de dados complexos de saúde, incluindo dados biométricos, geográficos, históricos, entre outros, poder-se-á identificar padrões e relações. A IA para ações em saúde pode transformar o campo da medicina, gerando melhores resultados para os pacientes e um sistema de saúde mais eficiente e eficaz.

ASPECTOS ÉTICOS NO USO DE IA NA MEDICINA

Para evitar problemas éticos decorrentes do uso da inteligência artificial na medicina, é imperativo estabelecer diretrizes claras e robustas que garantam transparência e confiabilidade nas decisões assistidas por IA. A essência desta abordagem reside na obrigação de tornar explícito o papel dos sistemas de IA na tomada de decisões clínicas. Isso não apenas mitigaria a desconfiança entre médicos, pesquisadores e pacientes, mas também fortaleceria a relação entre pacientes e o sistema de saúde. Uma comunicação clara sobre como e por que uma decisão foi apoiada por IA é crucial para manter a confiança e o entendimento mútuo.

Outro aspecto ético significativo é a responsabilidade pela guarda dos dados privados do paciente e pelos resultados gerados pelo uso da IA. Quando a IA é empregada

por profissionais de saúde, a responsabilidade tende a recair sobre eles, tanto pela segurança dos dados quanto pelos possíveis erros de diagnóstico ou tratamento. Contudo, a situação se complica quando os pacientes utilizam diretamente essas ferramentas de IA para autodiagnóstico ou tratamento. Mesmo que indiretamente, a disponibilização para o público geral de programas de IA com fins médicos pode prejudicar a interação profissional-paciente, com consequências imprevisíveis. Se um sistema de busca como o Google já espalhou desinformação e causou danos a muitos indivíduos, imagine-se o que pode acontecer com a democratização de modelos de IA que não são regulados ou testados devidamente? Nesses casos, a atribuição de responsabilidade é menos clara, levantando questões sobre a responsabilidade civil por danos ou diagnósticos incorretos. Por isso, a democratização (no sentido de ampla disponibilização) de programas de IA médica ao público pode prejudicar a relação profissional-paciente, com potenciais riscos de desinformação semelhantes aos observados em plataformas de busca na internet.

A questão da responsabilidade é central no debate ético sobre a IA na medicina. A ambiguidade sobre quem carrega a responsabilidade por erros ou danos – sejam eles os desenvolvedores de IA, os profissionais de saúde que a utilizam ou as próprias instituições – exige a criação de regulamentos específicos. Essas regras devem definir limites claros e atribuições de responsabilidade para proteger os pacientes contra prejuízos e violações de privacidade. É essencial que tais regulamentos sejam acompanhados de mecanismos de fiscalização efetivos para garantir sua aplicação. Acelerar o processo de regulamentação para

acompanhar o ritmo frenético de desenvolvimento da IA na medicina se torna, então, uma prioridade. Isso exige uma abordagem proativa e colaborativa que envolva legisladores, profissionais de saúde, desenvolvedores de IA, pacientes e a sociedade em geral. Uma estratégia eficaz pode incluir o estabelecimento de comitês regulatórios especializados em IA na saúde, que funcionem tanto em nível nacional quanto internacional, para garantir a criação de normas coerentes e aplicáveis globalmente.

O impacto da IA generativa e outras tecnologias de *machine learning* na medicina requer uma reavaliação dos códigos

éticos tradicionais. As escolas de medicina e outras instituições educacionais serão desafiadas a atualizar seus currículos e diretrizes éticas para abordar as nuances específicas trazidas pela IA. Esse processo inclui o enfrentamento do desafio das “caixas-pretas” em IA, em que os processos de decisão de modelos de *machine learning* são desconhecidos e não facilmente interpretáveis por humanos. Para superar isso, é crucial desenvolver e implementar técnicas de IA explicável, que tornem os processos de tomada de decisão da IA transparentes e compreensíveis, tanto para os profissionais de saúde quanto para os pacientes.

REFERÊNCIAS

- BHANDARI, M.; ZEFFIRO, T.; REDDIBOINA, M. “Artificial intelligence and robotic surgery: current perspective and future directions”. *Current Opinion in Urology*, v. 30, n. 1, jan./2020, pp. 48-54.
- HASHIMOTO, D. A. et al. “Artificial intelligence in surgery: promises and perils”. *Annals of Surgery*, v. 268, n. 1, jul./2018, pp. 70-6.
- KALA RANI, K. S. “Chapter 14: applications of quantum AI for healthcare”, in P. Raj et al. (eds.). *Quantum computing and artificial intelligence: training machine and deep learning algorithms on quantum computers*. Berlin/Boston, De Gruyter, 2023, pp. 271-88.
- KAUSHIK, R.; KANT, R.; CHRISTODOULIDES, M. “Artificial intelligence in accelerating vaccine development – Current and future perspectives”. *Frontiers in Bacteriology*, v. 2, 2023.
- KUMAR, Y. et al. “Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda”. *Ambient Intel Humanized Computing*, v. 14, n. 7, 2023, pp. 8.459-86.
- SIMOENS, S.; HUYS, I. “R&D costs of new medicines: a landscape analysis”. *Frontiers Medicine*, v. 8, 2021.

Inteligência artificial, direito e pesquisa jurídica

Marco Almada
Rafael A. F. Zanatta

resumo

O avanço da IA se reflete cada vez mais no direito, em variadas aplicações e técnicas. Nos últimos dez anos, diferentes projetos de uso de IA se institucionalizaram no sistema de justiça no Brasil. O presente artigo mapeia as implicações dessa transformação para a pesquisa em direito. Argumenta-se que, embora a IA jurídica tenha grande potencial, pode também levar a erros e até mesmo amplificar injustiças estruturais na sociedade. Por isso, o artigo identifica questões centrais para a discussão de quando, como e onde é desejável usar IA para a pesquisa jurídica.

Palavras-chave: IA e direito; tecnologia no Judiciário; pesquisa jurídica; aprendizado de máquina no direito.

abstract

The advancement of AI is increasingly reflected in law, in various applications and techniques. In the last ten years, different projects using AI have been institutionalized in the Brazilian judicial system. This article maps the implications of this transformation for legal research. It is argued that although legal AI has great potential, it can also lead to errors and even amplify structural injustices in society. In this context, the article identifies central questions for the discussion of when, how and where it is desirable to use AI for legal research.

Keywords: AI and law; technology in the Judiciary; legal research; machine learning in law.

A

relação entre inteligência artificial e o direito é mais antiga do que pode parecer à primeira vista. Já no século XVII, o filósofo e jurista alemão Gottfried Wilhelm Leibniz iniciou os esboços de uma primeira formalização de conceitos jurídicos (Artosi; Sartor, 2018), uma empreitada que serviu de inspiração para muitos esforços posteriores de automação do raciocínio jurídico (Sartor; Branting, 1998; Mumford; Atkinson; Bench-Capon, 2022). Com o advento da IA como disciplina técnica a partir dos anos 1940, este trabalho teórico logo ganhou uma dimensão prática. Acadêmicos e operadores do direito logo viram oportunidades para análises estatísticas do direito (Loevinger, 1949), iniciando uma tendência

aos estudos quantificados do direito que hoje é amplificada pela disponibilidade de técnicas de aprendizado de máquina (Dyevre, 2022). A IA no direito não é, portanto, uma novidade.

Mesmo tendo longos antecedentes históricos, o movimento brasileiro de fomento aos sistemas de IA no Judiciário é um fenômeno dos últimos dez anos (Costa Abreu; Silva, 2020; Moraes, 2023). Recentemente, chamou a atenção a notícia do Supremo Tribunal Federal de uma chamada pública, anunciada em novembro de 2023, para propostas de “modelos de inteligência artificial generativa, que pos-

MARCO ALMADA é pesquisador do Instituto Universitário Europeu (Florença, Itália).

RAFAEL A. F. ZANATTA é advogado, doutor pelo Instituto de Energia e Ambiente da USP e codiretor da Data Privacy Brasil.

sam dar suporte à prestação jurisdicional em processos de natureza pública” (STF, 2023). Vale destacar, no entanto, que as experiências inovadoras de uso de IA surgiram de necessidades práticas de diferentes unidades do Poder Judiciário.

Já em 2017, em Rondônia, o projeto Sinapses surgiu para solucionar um problema de automatização do processo de concessão de medicamentos, minimizando o esforço realizado por assessores em pesquisas e triagens. A instituição de uma equipe de inteligência artificial ocorreu para solucionar um problema concreto, porém as escolhas de *design* de tecnologia favoreceram sua utilização para além do Tribunal de Justiça de Rondônia. A equipe do Sinapses formulou um sistema de “microsserviços de IA”, que proporcionou o controle dos modelos, a gestão das versões e a rastreabilidade do treinamento¹.

Além do Sinapses, soluções de IA também se disseminaram como nos projetos ELIS (Pernambuco), Hércules (Alagoas), Horus (STJ) e Victor (STF), todos criados nos últimos anos. Apesar das diferentes técnicas, um elemento comum desses projetos é evitar trabalhos repetitivos de classificação e acesso a documentos, conquistar maior celeridade processual e efetivar o princípio de eficiência na prestação do serviço jurisdicional (Costa Abreu; Silva, 2020).

1 Como informa o CNJ, “uma vez encapsulados no Sinapses, os modelos podem ser servidos a qualquer sistema que necessite de uma resposta específica [...], gerando, assim, predição por meio de APIs RESTful” (CNJ, 2019, p. 14). O Sinapses tornou-se conhecido pela elegante escolha de *design*, que permite que o sistema cliente opere de forma independente do processo de construção dos modelos de IA. Isso facilitou sua disseminação, por times de cientistas de dados, em outras unidades do Judiciário.

Considerando a natureza multifacetada do presente dossiê sobre IA, apresentamos neste ensaio uma avaliação sobre como a crescente expansão dos sistemas de IA no sistema de justiça pode implicar mudanças significativas na pesquisa jurídica, que precisam ser encaradas tanto do ponto de vista das suas ambiguidades (facilitação do acesso *vs* opacidade dos sistemas de IA) quanto em seus impactos mais profundos com relação ao tensionamento entre o modo de pensamento jurídico e as ciências matemáticas e computacionais.

Aqui usamos o termo “pesquisa jurídica” em um sentido amplo para nos referirmos ao processo contínuo e dinâmico de investigação e análise de questões jurídicas com o objetivo de encontrar soluções, embasando decisões e contribuindo para o desenvolvimento do direito². Dado o caráter tecnológico do direito (Ferraz Jr., 2015), a pesquisa jurídica não se restringe às atividades da academia, mas analisa as práticas de pesquisa dos diversos profissionais dentro do sistema de justiça, incluindo advogados, promotores e juízes. Por isso, nosso enfoque de análise não se limita à pesquisa acadêmica em direito e aos debates a respeito do status epistêmico da produção de conhecimento jurídico dentro da academia³. Em vez disso, tratamos da questão mais ampla a respeito de como sistemas de IA cada vez mais surgem como intermediários das práticas de interpretação, aplicação e desenvolvimento do direito. Entender essa intermediação

2 A respeito do direito como prática social, ver: Hyden (2015).

3 Sobre esse tema, ver: Coyle; Pavlakos (2005).

é, portanto, essencial para compreender como a pesquisa jurídica dentro e fora da academia (e, portanto, o funcionamento da lei) é transformada pela adoção de tecnologias de IA.

APRENDIZADOS E VIRTUDES DA IA NA PESQUISA JURÍDICA

A adoção de ferramentas de IA no direito brasileiro é moldada por, pelo menos, dois fatores. De um lado, o vasto número de demandas judiciais a serem tratadas gera uma busca por soluções que permitam o acompanhamento dos processos e sua condução em tempo hábil. Segundo os dados do Conselho Nacional de Justiça apresentados no relatório “Inteligência Artificial no Poder Judiciário Brasileiro” (CNJ, 2019), existiam, em 2018, 78 milhões de processos judiciais em trâmite, sendo que cada um dos mais de 18 mil juízes decide, em média, 1.800 processos por ano. Diante de tal volume de trabalho, o potencial de automação associado à IA motivou a adoção de técnicas de IA por entes como escritórios de advocacia, a Advocacia-Geral da União (Granada, 2023) e o próprio CNJ (Salomão; Tauk, 2023).

Uma vez tomada a decisão pela automação, a escala das demandas judiciais cria condições propícias para o treinamento de sistemas de aprendizado de máquina. Tal volume de dados não é um desenvolvimento recente, e motivou diversas iniciativas de digitalização dos conteúdos de processos judiciais, dos “metadados” que os descrevem e da própria tramitação de tais processos. E, uma vez

que tais informações já se encontravam disponíveis em formato digital, diversos instrumentos regulatórios e decisões judiciais garantiram aos jurisdicionados um acesso à informação que encontra paralelo em poucos países do mundo (Trevelin et al., 2020). Um desenvolvedor que deseje criar ferramentas de IA para o Judiciário brasileiro tem, portanto, acesso a grandes bases de dados históricas e contemporâneas a respeito dos processos. Não é à toa que hoje o Brasil conta com um grande número de iniciativas de IA judicial, tanto desenvolvidas pela iniciativa privada (AB2L, 2024) quanto pelos próprios tribunais (Salomão; Tauk, 2023).

Quais são os papéis que a IA desempenha nos processos de pesquisa jurídica? De forma geral, essas tecnologias são usadas com o intuito de facilitar a interação de juristas e não juristas com as regras aplicáveis (McCarthy, 1990). Tal facilitação pode tomar múltiplas formas. Em alguns casos, ela envolve o acesso a conteúdos jurídicos já existentes por meio do refinamento de técnicas de busca e de extração de informação (Dyevre, 2021) ou de organização de acervos de conteúdos jurídicos (Zahir, 2023). Em um passo além, técnicas de IA podem ser utilizadas para a sumarização de informação presente nos autos de um processo (Granada, 2023) ou para a elaboração automatizada de explicações de conceitos jurídicos para o público leigo (Dahl et al., 2024, pp. 4-7). Tais abordagens, quando bem-sucedidas, prometem reduzir a carga cognitiva envolvida em lidar com questões judiciais.

Outras ferramentas vão além de usos pretensamente descritivos. Por exemplo,

algumas linhas de pesquisa propõem que *large language models* (LLMs) podem ser utilizados para estimar o significado de cláusulas ambíguas em contratos (Hoffmann; Arbel, 2024). Além disso, há uma ampla linha de pesquisas que busca prever como um caso será decidido por um juiz com base em informações processuais (Chalkidis et al., 2019). Por fim, há também propostas de utilizar IA na geração dos resultados da pesquisa jurídica, por exemplo, através da proposição de soluções alternativas para litígios (Ibodullaev, 2022). Muitas dessas pesquisas ainda não estão incorporadas aos fluxos de trabalho jurídicos, mas o potencial de atingir tais resultados é muitas vezes invocado como uma justificativa para o uso da IA⁴.

Além das aplicações voltadas aos pesquisadores jurídicos tradicionais, a IA também tem o potencial de envolver novos atores no processo de pesquisa. A crescente disponibilidade de assistentes inteligentes e chatbots jurídicos promete reduzir as barreiras de acesso à pesquisa jurídica, na medida em que permitam aos não juristas obter explicações de conceitos técnicos e do que acontece em um processo. Caso tais promessas se materializem, elas contribuirão não só para facilitar a pesquisa jurídica dos especialistas na área, como para tornar o sistema jurídico legível para o público em geral.

À medida que o sistema jurídico brasileiro abraça a era da inteligência artificial, há a expectativa de que a

pesquisa jurídica seja transformada, tornando-se mais eficiente e acessível para especialistas e o público em geral. O projeto Hércules, por exemplo, prevê classificações de tipos de petições e documentos jurídicos e sugere movimentações com base em padrões, o que gera ganhos de eficiência, celeridade processual e benefícios para a comunidade. Propostas como a do Supremo Tribunal Federal de usar IA generativa para criar sumários automatizados ou sintetizar o que é decidido pela Corte podem ter implicações cidadãs significativas. Sistemas de IA podem gerar análises automatizadas de casos julgados em um mesmo dia, gerando informativos por WhatsApp por exemplo. Por estarem em fases incipientes, podemos apenas especular sobre seus potenciais efeitos sociais. Todavia, as condições atuais sugerem que eventuais benefícios democráticos decorrentes do uso da IA serão incidentais⁵, seguindo como consequência de outros objetivos perseguidos pelo Judiciário e pelos atores privados que desenvolvem IA para a pesquisa jurídica.

Analisando as movimentações do Conselho Nacional de Justiça e das Cortes de cúpula no Brasil nos últimos anos, nos parece claro que o uso de sistemas de IA é um componente de uma nova fase de políticas judiciárias. A orientação central da maioria dos projetos de IA está em

4 Ou para sua proibição: a última reforma do Judiciário francês incluiu um dispositivo que proíbe a predição individualizada de como juízes decidem casos (Langford; Madsen, 2019).

5 Como argumentou Juliano Maranhão, em palestra ao STF em 2022, intitulada "Inteligência artificial no Poder Judiciário", os data systems no direito ajudam a estimar tendências e definir estratégias de acordo, buscar fontes jurídicas (leis, precedentes, doutrina, documentos do Judiciário), organizar informação relevante para análise, classificar textos e documentos jurídicos e acompanhar andamento processual extraindo informação relevante.

busca de solução de problemas concretos de sobrecarga de trabalho, dificuldade de organização de informações muito pulverizadas em documentos jurídicos (petições, documentos juntados como provas, decisões etc.) e ineficiência na prestação do serviço jurisdicional (Costa Abreu; Silva, 2020). A adoção de sistemas de IA surge, assim, como um desdobramento de outras frentes de política digital no Judiciário, como transformação digital, organização dos processos eletrônicos e uniformização da estratégia de usos de tecnologias pelo sistema de justiça. Embora tais processos tenham seus limites⁶, eles contribuem para que o uso de dados no contexto judicial brasileiro seja marcado por um maior grau de interoperabilidade e compartilhamento de dados do que é o caso em outros domínios de uso da IA, ou mesmo em iniciativas de IA jurídica em outras partes do mundo⁷. Há, também, uma continuidade no sentido de aprendizado institucional com políticas do passado, como aquelas sobre processo eletrônico judicial⁸. Não se trata, portanto, de um evento isolado, mas de uma peça em mudanças sistêmi-

cas que prometem alterar a forma como a pesquisa jurídica é feita no Brasil.

DESAFIOS E AMBIGUIDADES DO PROCESSO DE PESQUISA JURÍDICA MOVIDA A IA

As promessas associadas à IA na pesquisa jurídica devem, contudo, ser tomadas com um certo ceticismo. Um motivo para tal é a chamada *falácia da funcionalidade* (Raji et al., 2022): muitas vezes, toma-se como dado que um sistema funciona como prometido em sua especificação, mas nem sempre isso é verdade, seja porque as promessas foram exageradas ou distorcidas, seja por problemas no projeto do sistema. Aplicações da IA no campo jurídico não estão imunes a esses problemas.

Recentes estudos a respeito do uso de LLMs no direito, por exemplo, apontam que sua utilidade prática é reduzida pela alta taxa de “alucinações”, isto é, de respostas que não são consistentes com fatos a respeito do caso ou do sistema jurídico (Feijó; Moreira, 2023). A frequência de tais alucinações está ligada a diversos fatores, como a complexidade da tarefa jurídica, o tipo de modelo usado e a disponibilidade de informação a respeito de uma jurisdição (Dahl et al., 2024). À primeira vista, esse último fator encorajaria o uso de tais modelos no direito brasileiro, dado o acesso aos dados descrito acima. Contudo, os LLMs atualmente disponíveis para o público ainda geram resultados consideravelmente distorcidos, o que pode induzir juízes, advogados e outros profissionais ao erro e fazer com que os sistemas não sejam uma fonte confiável para quem não tem o conhecimento

6 Por exemplo, a linguagem de marcação LexML, desenvolvida como parte das iniciativas de governo eletrônico brasileiro, não prosperou como padrão de representação de dados: LexML, Portal do Projeto LexML. Disponível em: <https://projeto.lexml.gov.br/>. Acesso em: 28/fev./2024.

7 O que pode ser visto ao comparar a situação acima com o quadro fragmentado de iniciativas de IA jurídica na Europa: Cepej, Dynamic database of European judicial systems. Disponível em: <https://www.coe.int/en/web/cepej/cepej-stat>. Acesso em: 28/fev./2024.

8 O claro investimento em uma lógica aberta de API e construções modulares pode ser visto como aprendizado com problemas do passado em relação a investimentos em tecnologia que não possuíam interoperabilidade, hoje tido como princípio central para o Conselho Nacional de Justiça (Ikawura, 2020).

técnico para fazer uma avaliação crítica dos resultados gerados.

Mesmo quando os sistemas produzem resultados tecnicamente corretos, eles ainda podem dar origem a uma série de riscos. Um deles é a criação de uma “monocultura jurídica” (Dahl et al., 2024, p. 2): se todos utilizam os mesmos sistemas para interpretar o direito, os resultados da interpretação serão sempre similares. Se, de um lado, isso contribui para a segurança jurídica e previsibilidade da lei (Hoffman; Arbel, 2024), de outro, acaba por fechar as possibilidades de mudança na interpretação do direito, que são justamente o que permite que ele se adapte às demandas sociais. Tornar o direito inteiramente previsível pode transformá-lo, ao mesmo tempo, num fóssil que não cumpre seu papel de produzir justiça e proteger os direitos individuais e coletivos.

Além disso, a complexidade técnica dos sistemas de IA e das infraestruturas de dados e processamento necessárias para seu uso faz com que os mercados para ferramentas tecnológicas sejam desproporcionalmente dependentes de um punhado de atores econômicos (ver, e.g., Almada et al., 2024). A pesquisa intermediada por IA também se torna suscetível a influências de atores políticos que possam controlar o acesso aos dados e o uso das ferramentas. No Brasil, por exemplo, muito do acesso a dados judiciais é moldado por normas estabelecidas pelo CNJ, que define formatos de informação e diretrizes para o acesso programático às informações disponibilizadas pelos tribunais. Tais atores se tornam intermediários da pesquisa jurídica baseada em IA e suas decisões técnicas e de negócio podem afetar a qualidade, o funcionamento, os custos e mesmo a viabilidade de diferentes ferramentas.

Essa mesma complexidade contribui para que o uso da IA torne a pesquisa jurídica mais opaca. Sistemas de IA modernos são objetos técnicos complexos, cujo funcionamento não é acessível para leigos. Algumas técnicas buscam tornar o funcionamento desses sistemas mais inteligíveis, mas, em contextos jurídicos, tais abordagens são passíveis de manipulação pelos desenvolvedores do sistema, que controlam os parâmetros e técnicas usadas para gerar transparência (ver, por exemplo, Bordt et al., 2022). Dessa forma, tanto leigos quanto juristas carecem dos meios para avaliar os processos de tomada de decisão dos sistemas de IA que utilizam na pesquisa jurídica.

A opacidade dos sistemas de IA cria uma barreira ao escrutínio dos resultados que eles geram na pesquisa jurídica. Diante das dificuldades técnicas em saber o que se passa dentro de um sistema, tanto leigos como juristas muitas vezes acabam por simplesmente tomar por certo que os sistemas atuam como esperado, caindo na falácia da funcionalidade descrita acima. Mesmo que um indivíduo tenha tempo e interesse para entender como uma determinada resposta da IA é produzida, a complexidade técnica e as barreiras jurídicas criadas pela proteção da propriedade intelectual dos fornecedores das ferramentas tecnológicas reduzem a visibilidade dos mecanismos pelos quais a IA produz suas decisões. A integração da IA no processo de pesquisa jurídica pode, portanto, torná-lo uma caixa-preta, desconectando o pesquisador das práticas de produção de conhecimento e decisões jurídicas.

Por fim, é importante ressaltar uma série de obstáculos técnicos ao uso da IA no âmbito jurídico. Como aponta a falácia da funcionalidade, nem sempre os sistemas

de IA operam como esperado, e seu uso adequado exige uma série de pré-condições técnicas. Um recente trabalho a respeito do uso de técnicas de IA generativa no direito (Cooper et al., 2023), por exemplo, aponta lacunas no entendimento de pesquisadores a respeito das capacidades desse tipo de sistema e das peculiaridades do domínio jurídico. Tal falta de conhecimento contribui para a opacidade dos sistemas de IA, mas também cria problemas para o projeto de sistemas aplicados ao direito.

A esses problemas epistêmicos, somam-se problemas de ordem material. A concentração de poder apontada acima faz com que a infraestrutura dos sistemas de IA usados na pesquisa jurídica seja vulnerável a falhas, brechas de segurança, ou mesmo ao arbítrio de alguns provedores. Aqui, nos referimos não só às empresas – ou entes públicos – que criam os sistemas propriamente ditos, mas também àqueles que fornecem os servidores de nuvem nos quais os dados são armazenados e o processamento ocorre, bem como o hardware utilizado para tais operações⁹. Os problemas decorrentes dessa concentração não são exclusivos da pesquisa jurídica, mas permeiam o uso de IA nesse domínio.

Não há solução simples para quaisquer dos desafios elencados acima. Ademais, todos eles tornam o processo de pesquisa jurídica dependente das ações de desenvolvedores de *software*, definidores de normas técnicas e outros atores envolvidos no projeto e comercialização de sistemas de IA. É necessário, portanto, investigar como o

Estado de direito pode ser enfraquecido à medida que a IA se vai enraizando em uma sociedade tecnocrática, em razão da diminuição da capacidade de compreensão e oferecimento de razões públicas para determinadas escolhas (Greenstein, 2022). O investimento em equipes de cientistas de dados e times de IA no sistema de justiça no Brasil precisa estar orientado não apenas para os problemas internos de solução de questões matemáticas, mas também para a orientação pública de por que determinados sistemas de IA são benéficos para a sociedade como um todo. Nesse sentido, é preciso que os times de IA no sistema de justiça sejam capazes de apresentar explicações e razões na esfera pública, como defende Stuart Russell (2019).

QUESTÕES TEÓRICAS EMERGENTES NO AVANÇO DA IA NA PRÁTICA JURÍDICA

Algumas das preocupações atuais com o avanço da IA remontam aos debates da cibernética jurídica da década de 1960. O jurista italiano Vittorio Frosini era um entusiasta da “aplicação da cibernética no setor administrativo público” (Frosini, 1973, p. 103), porém também se preocupava com os limites da ressignificação das práticas jurídicas em razão do domínio da lógica, da calculabilidade, da automação e da predição no direito.

O que inquietava Frosini, há mais de 50 anos, era a limitação da experiência jurídica a uma estrutura formal, lógica e matemática; uma espécie de raciocínio perfeitamente objetivo e “totalmente tecnicizado” (Frosini, 2023). Prevendo a ascensão dos computadores e dos sistemas de inteligência artificial,

9 A respeito dos diversos atores envolvidos nas cadeias de suprimento de IA, ver: Cobbe; Veale; Singh (2023).

Frosini defendeu que juristas tivessem capacidade de compreensão, inovação e operação em sinergia com as máquinas. Para ele, os computadores podem e devem participar da aplicação do direito no sistema de justiça. Porém, a automação nunca deve ser total. A transformação do sistema de justiça pelos computadores e pelas tecnologias baseadas em IA deve estar sujeita ao controle, responsabilidade e avaliação crítica dos juristas¹⁰.

Para alguns, essa transformação na forma da pesquisa jurídica não seria algo negativo. Introduzir a IA como um intermediário na interpretação e aplicação do direito seria uma forma de agilizar processos (Someson; Tauk, 2023), resolver disputas interpretativas (Hoffman; Arbel, 2024) e de lastrear a pesquisa jurídica em técnicas de análise de dados que são cada vez mais salientes nas ciências sociais (Dyevre, 2021), dentre outras potenciais vantagens. Mesmo que tais características sejam desejáveis (ao menos até certo ponto), entendemos que elas devem ser balanceadas com os outros valores que a pesquisa jurídica deve promover, como a busca pela justiça, a confiabilidade do sistema jurídico e a flexibilidade para atender às diferentes demandas que surgem de sociedades em constante mudança. Tal conciliação, entendemos, exige uma atenção renovada à teoria do direito e ao “direito computacional” (Hildebrandt, 2018).

Tal olhar teórico é necessário porque a pesquisa jurídica empírica não tem por

objeto fatos sobre o mundo. Em vez disso, seu objeto são normas jurídicas, que geram expectativas contrafactuais: uma lei continua a ser lei ainda que ela seja desrespeitada em casos particulares. Uma visão do direito que se baseia em análises estatísticas encontra uma série de obstáculos, dos quais introduziremos dois.

O primeiro é que modelos baseados em estatísticas resultam em uma distorção do funcionamento do direito (Diver; McBride, 2022). Uma exigência do funcionamento do direito, que inclusive tem força de lei, é que os casos devem ser tratados à luz de suas características específicas. Tal individualização dos casos vai no sentido oposto àquele de técnicas baseadas em aprendizado de máquina, que buscam justamente tratar de casos individuais a partir de padrões estatísticos. Assim, o uso de aprendizado de máquina para determinadas funções na pesquisa jurídica exige um cuidado para que seja suprida a individualização que essas técnicas não provêm. A lógica de detecção de padrões em vastas bases de dados e correlações por técnicas de aprendizado de máquinas (Hildebrandt, 2008) não necessariamente se apresenta como melhor estratégia para a decidibilidade de casos concretos. Este é um ponto que merece bastante atenção da comunidade científica.

Por último, mas não menos importante, o desenvolvimento de sistemas de IA jurídica deve estar atento às diferenças entre os processos de aprendizado de máquina e a forma como humanos raciocinam a respeito de questões jurídicas. Em particular, tanto humanos como máquinas podem ajustar seu aprendizado conforme recebem novas informações, mas o fazem de formas diferentes (Binns, 2022, p. 6): para um sistema de

10 Como afirmou Fabiano Hartmann Peixoto, coordenador do projeto Victor (STF), em seminário do Instituto de Estudos Avançados da USP, um jurista não precisa ter conhecimentos avançados de programação e ciência da computação para se envolver nos usos de IA no sistema de justiça, mas precisa ser capaz de ter uma discussão metodológica com os times de IA.

aprendizado de máquina, uma nova decisão judicial é um novo ponto a ser considerado, mas um jurista humano sabe que uma única decisão por vezes pode levar a uma reconfiguração drástica do sistema jurídico. Uma sentença de que uma determinada lei é inconstitucional faz com que quaisquer julgamentos futuros não possam aplicar as regras que antes valiam. Um sistema de IA para a pesquisa jurídica não pode se basear, portanto, em um modelo de evolução gradual, mas deve também ter um processo de evolução que englobe tais rupturas¹¹.

Tomemos como exemplo o problema das decisões sobre racismo nos EUA. Por décadas, decisões judiciais nos EUA reforçaram as teses racistas de separação entre brancos e negros, a separação de espaços específicos para negros e a competência dos estados para decidir elementos de organização social, incluindo tais componentes sobre acesso a bens e serviços. Imaginemos, hipoteticamente, que uma IA tenha acompanhado todo esse processo, entre o final do século XIX até a década de 1950. Como um sistema de IA formularia o impacto de uma decisão como “Brown v. Board Education”, que declarou inconstitucionais várias das leis que sustentavam esse regime? Um dos possíveis focos de análise para pesquisas futuras é a compreensão do papel dos “litígios estruturais” ou “litígios estratégicos” e

sua relação com a organização do conhecimento jurídico no sistema de justiça, tendo em mente a estrutura matemática dessas técnicas de compreensão de padrões e os modelos de representação que as sustentam. Isso envolve uma sofisticada discussão entre direito e ciência da computação para preparação dos dados, rotulagem e desenho de um sistema capaz de entender em que condições uma decisão está apta a “nadar contra a corrente”.

Esses e outros fatores não são, necessariamente, obstáculos intransponíveis à adoção da IA na pesquisa jurídica. As (potenciais) vantagens de uma determinada aplicação podem ser tão grandes que motivem o legislador a alterar a lei para que ela se torne mais mecânica e mais amena à automação. Ou, o que nos parece mais provável, os projetistas de sistemas de IA jurídica podem incorporar medidas que reduzam potenciais impactos adversos da IA na pesquisa jurídica. Mas, para que tais respostas sejam bem-sucedidas, não basta o entendimento das leis. É necessário um engajamento mais profundo com teorias a respeito do que é e do que deveria ser o direito. De forma espelhada, é preciso também um engajamento dos juristas com os elementos metodológicos da ciência de IA e um letramento em lógica e matemática.

CONCLUSÃO

De um ponto de vista tecnológico, muito mudou desde as reflexões de Frosini na década de 1960 sobre cibernética e sistema de justiça. Como reconhecido por Giovanni Sartor e Karl Branting (1998), o progresso no processamento da lingua-

¹¹ Como notado pelo cientista da computação Trevor Bench-Capon, um dos pioneiros em modelagem matemática de argumentação jurídica para sistemas de IA, o aprendizado de máquina é retrospectivo (treinado em decisões passadas), enquanto a jurisprudência é prospectiva (destinada a influenciar decisões futuras): “Ao decidir um caso não estamos a descobrir algo comum aos casos anteriores, estamos a criar uma regra para decidir um caso particular, e que se destina a condicionar casos futuros” (Bench-Capon, 2020).

gem natural, raciocínio qualitativo e redes neurais alargou a gama de experiências e comportamentos humanos abordados pela IA. No campo jurídico, o uso da IA é bem mais diverso e complexo que a automação de procedimentos jurídicos rotineiros, como procedimentos administrativos e a aplicação direta de regras claras a fatos não controversos, como na década de 1960. A utilização de sistemas de IA para questões mais complexas, como IA para apoio à tomada de decisão judicial ou usos de IA generativas para sínteses de decisões sobre fatos e questões jurídicas relevantes, tem fomentado grandes debates na intersecção entre direito e ciência da computação. Não só para Sartor, mas também para Bench-Capon, algoritmos de reconhecimento de padrões são incapazes de modelar com precisão a tomada de decisões judiciais, pois o sistema de justiça sempre enfrentará casos em que as relações entre fatos e regras são controversas, exigindo uma reinterpretação que modifica decisões futuras.

Tarefas computacionais que antes pareciam impossíveis hoje são parte do nosso dia a dia, e isso se reflete nas capacidades e expectativas para a IA no direito. Ainda

assim, muitas das questões discutidas no passado encontram eco na literatura contemporânea que examinamos acima. Para entendermos o que é a pesquisa jurídica contemporânea, precisamos entender o que muda – e o que permanece igual – à luz das diferentes escolhas tecnológicas.

Tal entendimento exige diversas linhas de investigação. De um ponto de vista teórico, há a necessidade de maior estudo a respeito das diferenças entre a pesquisa jurídica e as inferências sobre fatos que são o objeto da maior parte das aplicações de IA. De um ponto de vista sociojurídico, futuras investigações podem trazer mais informações tanto sobre os processos que levam os tribunais a adotar sistemas de IA quanto sobre o impacto de diferentes formas de automação para os juristas e para o público leigo. Por fim, é necessário também um maior estudo dos desafios tecnológicos envolvidos na construção de sistemas de IA voltados a tarefas jurídicas. Estas linhas de investigação certamente não cobrem todo o impacto da IA no direito. Mas, ao formulá-las, nossa intenção é fornecer um ponto de partida para a análise da IA no contexto da pesquisa jurídica.

REFERÊNCIAS

- AB2L. "Radar de lawtechs e legal techs". 2023. Disponível em: <https://ab2l.org.br/ecossistema/radar-de-lawtechs-e-legaltechs/>. Acesso em: 2/fev./2024.
- ALMADA, M.; MARANHÃO, J.; SARTOR, G. "Competition in and through artificial intelligence", in P. L. Parcu, M. A. Rossi, M. Botta (orgs.). *Research handbook on competition and technology*. Cheltenham, Edward Elgar Publishing, 2024 (no prelo).
- ARTOSI, A.; SARTOR, G. "Leibniz as jurists", in M. R. Antognazza (ed.). *Oxford Handbook of Leibniz*. Nova York, Oxford University Press, 2018, pp. 641-63.
- BENCH-CAPON, T. "The need for good old fashioned AI and law". *Jusletter-IT*, 2020.
- BINNS, R. "Analogies and disanalogies between machine-driven and human-driven legal judgement". *Journal of Cross-disciplinary Research in Computational Law*, v. 1, n. 1, 2022.
- BORDT, S. et al. "Post-hoc explanations fail to achieve their purpose in adversarial contexts". *FACCT'22*. Seul/Nova York, ACM, 2022, pp. 891-905.
- CHALKIDIS, I. et al. "Paragraph-level rationale extraction through regularization: a case study on European Court of Human Rights cases". *NAACL-HLT 2021*. [S.l.], ACL, 2021, pp. 226-41.
- CNJ. *Inteligência artificial no Poder Judiciário brasileiro*. Brasília, Conselho Nacional de Justiça, 2019.
- COBBE, J.; VEALE, M.; SINGH, J. "Understanding accountability in algorithmic supply chains". *2023 ACM Conference on Fairness, Accountability, and Transparency*. Chicago, ACM, 2023, pp. 1.186-97.
- COOPER, A. F. et al. "Report of the 1st Workshop on Generative AI and Law". arXiv 2311.04577v3 [cs.CY], 2023.
- COSTA ABREU, M.; SILVA, B. "A critical analysis of 'Law 4.0': the use of automation and artificial intelligence and their impact on the judicial landscape of Brazil". *Revista de Direitos Fundamentais e Tributação*, v. 1, n. 3, 2020, pp. 1-16.
- COYLE, S.; PAVLAKOS, G. (orgs.). *Jurisprudence or legal science? A debate about the nature of legal theory*. Oxford, Hart Publishing, 2005.
- DIVER, L.; MCBRIDE, P. "Argument by numbers: the normative impact of statistical legal tech". *Communitas: Théories et Pratiques de la Normativité*, v. 3, n. 1, 2022.
- DYEVRE, A. "Text-mining for lawyers: how machine learning techniques can advance our understanding of legal discourse". *Erasmus Law Review*, v. 14, n. 1, 2021, pp. 7-23.
- FEIJO, D. de V.; MOREIRA, V. P. "Improving abstractive summarization of legal rulings through textual entailment". *Artificial Intelligence and Law*, v. 31, n. 1, 2023, pp. 91-113.
- FERRAZ JR., T. S. *Função social da dogmática jurídica*. 2ª ed. Rio de Janeiro, Atlas, 2015.
- FROSINI, V. "L'automazione elettronica nella giurisprudenza e nell'amministrazione pubblica". *Bollettino Bibliografico d'Informatica Generale e Applicata al Diritto*, v. 2, n. 3-4, 1973, pp. 101-4.
- FROSINI, V. *Cibernetica, Diritto e Società*. Introduzioni di Giovanni Sartor. Roma, Roma Tre Press, 2023.
- GRANADA, A. *What's left of law? Quantifying and automating litigation practice*. Dissertação de mestrado. São Paulo, Faculdade de Direito da USP, 2023.
- GREENSTEIN, S. "Preserving the rule of law in the era of artificial intelligence (AI)". *Artificial Intelligence and Law*, v. 30, n. 3, 2022, pp. 291-323.

- HILDEBRANDT, M. "Defining profiling: a new type of knowledge?", in *Profiling the European citizen: Cross-disciplinary perspectives*. Dordrecht, Springer Netherlands, 2008, pp. 17-45.
- HILDEBRANDT, M. "Law as computation in the era of artificial legal intelligence: speaking law to the power of statistics". *University of Toronto Law Journal*, v. 68, n. supplement 1, 2018, pp. 12-35.
- HOFFMAN, D. A.; ARBEL, Y. A. "Generative interpretation". *New York University Law Review*, v. 99, 2024 (no prelo).
- HYDÉN, H. "Perspectivas em sociologia do direito", in *Sociology of law on the move*. Canoas, UnilaSalle, 2015, pp. 67-92.
- IBODULLAEV, S. "Digitalization of the Civil Justice System: the experience of the Republic of Colombia". *International Bulletin of Applied Science and Technology*, v. 2, n. 11, 2022, pp. 88-94.
- IWAKURA, C. R. *Princípio da interoperabilidade: acesso à justiça e processo eletrônico*. São Paulo, Editora Dialética, 2020.
- LANGFORD, M.; MADSEN, M. R. "France criminalises research on judges". *Verfassungsblog*, 22/jun./2019. Disponível em: <https://verfassungsblog.de/france-criminalises-research-on-judges/>. Acesso em: 2/fev./2024.
- LOEVINGER, L. "Jurimetrics: the next step forward". *Minnesota Law Review*, v. 33, n. 5, 1949, pp. 455-92.
- MCCARTY, L. T. "Artificial intelligence and law: how to get there from here". *Ratio Juris*, v. 3, n. 2, 1990, pp. 189-200.
- MORAIS, G. R. *Pressupostos e diretrizes para a aplicação do processamento de linguagem natural à recuperação de informações jurisprudenciais: uma aplicação da AI4SG*. São Paulo, Editora Dialética, 2023.
- MUMFORD, J.; ATKINSON, K.; BENCH-CAPON, T. "Reasoning with legal cases: a hybrid ADF-ML approach". *IOS Press*, 2022, pp. 93-102.
- RAJI, I. D. et al. "The fallacy of AI functionality", in *FACt '22*. Seul/Nova York, ACM, 2022.
- RUSSELL, S. *Human compatible: artificial intelligence and the problem of control*. Londres, Penguin, 2019.
- SARTOR, G.; BRANTING, L. K. "Introduction: judicial applications of artificial intelligence", in *Judicial applications of artificial intelligence*. Dordrecht, Springer Netherlands, 1998, pp. 1-6.
- SALOMÃO, L. F.; TAUKE, C. S. (orgs.). *Inteligência artificial: tecnologia aplicada à gestão dos conflitos no âmbito do Poder Judiciário brasileiro*. 3ª ed. Rio de Janeiro, FGV, 2023.
- STF. Edital de chamamento público 1/2023: Soluções de IA generativa para sumarização em processos judiciais. Processo SEI 009336/2023. Brasília, STF, 2023. Disponível em: <https://www.stf.jus.br/arquivo/cms/noticiaNoticiaStf/anexo/ChamamentoPblicoASTF.pdf>.
- TREVELIN, B. et al. "Lawgorithm". *Acesso a Dados de Processos Judiciais no Brasil*. São Paulo, Lawgorithm, 2020. Disponível em: <https://lawgorithm.com.br/acesso-a-dados-de-processos-judiciais-no-brasil/>. Acesso em: 25/mai./2022.
- ZAHIR, J. "Prediction of court decision from Arabic documents using deep learning". *Expert Systems*, 2023.



**Inteligência artificial e internet:
um olhar sobre o conteúdo
de usuários e a sua moderação**

Francisco Brito Cruz

resumo

O artigo organiza o papel que a IA tem e pode ter sobre a circulação de conteúdos na internet. À medida que a produção de conteúdo gerado por usuários atinge volumes sem precedentes, a IA emerge como uma ferramenta indispensável para criar, filtrar, organizar e avaliar a vasta quantidade de informações. De sua parte, a emergência contínua dessa tecnologia traz consigo desafios significativos, incluindo o risco de perpetuar vieses e injustiças sistêmicas. O trabalho explica os diferentes eixos de impacto da IA e mergulha no seu uso para a moderação de conteúdos na internet, crucial para a segurança de usuários e proteção a direitos. Abordando riscos e formas de proteção e mitigação, expõe um panorama dos debates sobre o tema no momento em que os grandes modelos de linguagem passam a ser considerados como possíveis aliados de plataformas de internet em seu trabalho de aplicar regras para conteúdos gerados por usuários.

Palavras-chave: IA e internet; plataformas; moderação de conteúdo; curadoria nas plataformas; riscos da IA na internet.

abstract

The article outlines the role that AI currently plays in the dynamics of production, distribution and consumption of content on the internet. As user-generated content reaches unprecedented volume, AI emerges as an indispensable tool for creating, filtering, organizing, and evaluating vast amounts of information. However, the continuous emergence of this technology brings significant challenges, including the risk of perpetuating biases and systemic injustices. The paper explains the different impact axes of AI and delves into its use for content moderation on the internet, which is crucial for user safety and protection of human rights. Addressing risks and their mitigation, it presents an overview of the ongoing conversations on the subject, especially as large language models are considered potential allies for internet platforms in their efforts to enforce rules for user-generated content.

Keywords: AI and internet; platforms; content moderation; curation on platforms; AI and risks.

INTRODUÇÃO AO ESCOPO: INTERNET, PLATAFORMAS E CONTEÚDO GERADO POR USUÁRIOS

A inteligência artificial e internet são temas vastos por si sós. Ao apresentarem-se como tecnologias revolucionárias e repletas de ramificações e usos, escrever sobre o tema demanda escolhas e recortes. Na mesma interseção, por exemplo, é possível pesquisar o uso de tecnologias de aprendizagem de máquina para o oferecimento de uma miríade de serviços on-line pela internet, para a análise de dados pessoais e produção de perfis comportamentais para o direcionamento

de propaganda ou, ainda, para vulnerabilizar a segurança da informação de uma série de atividades que fazemos pela rede.

Neste texto o recorte é o uso de IA no contexto da produção, curadoria e, mais especialmente, moderação de conteúdos gerados por usuários da internet, processos que em geral têm ocorrido a partir de plataformas que possuam componentes “sociais” ou decorrentes da geração de conteúdo por terceiros na internet. A escolha por esse recorte não é por acaso: apesar de uma série de usos de IA desconectados das chamadas “plataformas” apresentarem questões sociais e econômicas críticas (como o impacto no mundo do trabalho, as questões socioambientais

FRANCISCO BRITO CRUZ é diretor executivo do InternetLab e autor de *Novo jogo, velhas regras* (Letramento).

e de matriz energética ou de segurança cibernética), a emergência e permanência da discussão sobre “desinformação”¹ no debate público desde meados da década de 2010 impõe especial atenção a de que forma tais serviços se propõem a serem espaços para produção de conteúdo gerado por usuários, como por vezes realizam a curadoria automatizada de enorme volume deste conteúdo, e, finalmente, como criam e aplicam regras sobre o que pode ou não ser postado ou circular – e como.

Assim, se considerarmos que a desordem informacional² é um problema relevante a ser discutido e atravessado pelos avanços nas tecnologias de IA, é fundamental compreendermos como no coração desse fenômeno estão presentes serviços de compartilhamento de conteúdos na internet que possuem como base o recebimento e organização de

informações geradas não por seus funcionários ou por pessoal contratado, mas sim por usuários alheios às estruturas dessas empresas.

Por mais que cada um desses serviços – ou “plataformas”, para facilitar – escolha estratégias diferentes de geração de receitas (frequentemente ligadas à venda de publicidade direcionada por dados pessoais extraídos dos mesmos usuários), o que os une é que se construíram atraindo usuários que neles podem ao mesmo tempo colocar sua atenção para consumir conteúdo alheio e produzir conteúdo próprio em diversas atividades de acesso à informação, comunicação e entretenimento. No jargão, são plataformas que acondicionam, curam e apresentam “conteúdo gerado por usuários” (*user-generated content*, ou UGC) em diferentes formatos multimídia.

Compreendido este contexto, três atividades emergem como fundamentais na configuração de plataformas “de conteúdo gerado por usuários”. Em primeiro lugar, a produção e postagem do conteúdo. Elas podem ser realizadas por indivíduos ou

1 Este é outro termo que merece comentário. Nos trabalhos desenvolvidos por este autor no âmbito do InternetLab, centro de pesquisa em direitos e tecnologias, *fake news* ou desinformação são alcunhas simplificadoras para um fenômeno que reflete a mudança de como as sociedades produzem, circulam e consomem informação política. Assim, o advento da internet como meio de comunicação política permitiu que qualquer indivíduo com conectividade se tornasse potencialmente um emissor de comunicação em massa, o que faz com que os novos conteúdos produzidos possivelmente se distanciem daqueles elaborados sob o imperativo do jornalismo profissional, de modo que as informações emitidas nem sempre têm como compromisso a busca por objetividade. Esse novo ambiente acaba por enfraquecer “ciclos de checagem da realidade”, impulsionados pelo jornalismo profissional, em favor de ciclos de “retroalimentação de propaganda”, nos quais a informação circula a partir de uma lógica político-partidária. A dinâmica da desinformação, assim, extrapola componentes de manipulação e de crise entre verdade e mentira, e envolve a própria forma como os indivíduos se relacionam com informação, em um processo de comunicação em rede no qual a autonomia desse indivíduo de produzir e compartilhar conteúdo ganha escala significativa (Brito Cruz, 2019).

2 Segundo Wardle e Derakshan (2017), o conceito de “desordem informacional” é o conjunto de três tipos de informação diferenciados de acordo com o seu grau de dano, sua inveracidade e a intenção presumida do seu remetente. São eles: (i) informação incorreta (*misinformation*); (ii) desinformação (*disinformation*); e (iii) conteúdo mal-informado (*malinformation*). Enquanto a informação incorreta corresponde ao fenômeno do compartilhamento de informações falsas sem a intenção de causar danos, a desinformação ocorre quando informações falsas são propagadas conscientemente para causar danos. Por sua vez, o conteúdo mal-informado contém informações genuínas, mas que são compartilhadas para causar danos, geralmente ao transferir para a esfera pública informações que deveriam permanecer privadas (por exemplo, vazamentos de e-mail, assédios on-line e discursos de ódio).

organizações de diversos tipos, de maneira mais ou menos organizada, e são sempre condicionadas pelos elementos arquitetônicos (ou *affordances*³) oferecidos pela plataforma. O exemplo mais clássico é o antigo Twitter, que apenas permitia postagens de texto que tivesse um número máximo de caracteres.

Em segundo, a atividade de curadoria e arquitetura de conteúdo gerado por usuários destina-se a organizar como tais peças multimídia irão viajar pela rede e serem entregues a diferentes audiências usuárias da plataforma, em diferentes contextos de uso. Na medida em que cada serviço possui uma estratégia de geração de receitas (que em geral passa por favorecer determinados usos e reter a atenção do público para revendê-la na forma de publicidade), cada um projeta a sua arquitetura a fim de proporcionar diferentes tipos de recomendação ou disposição gráfica de conteúdos gerados por terceiros. Essa distribuição pode ser simplesmente uma caixa de mensagens

que atualiza conforme as mais recentes forem recebidas ou enviadas – como no WhatsApp – ou um construto complexo de ferramentas de busca e recomendação ativa (ou seja, aquela em que o usuário não precisa dizer o que procura) que dispõe o conteúdo em diversos formatos – como no Instagram e YouTube.

Por fim, a atividade de moderação de conteúdo caracteriza-se quando a plataforma formula, edita, publica e aplica normas sobre o que pode ou não ser realizado por seus usuários. Conforme já explicitado em outro trabalho no âmbito do centro de pesquisa InternetLab,

“[...] essa é uma atividade de elaboração e aplicação de regras [públicas]. Com a finalidade de construir diferentes ambientes digitais propícios a interações sociais e geração de conteúdos por seus usuários, cada plataforma decide que tipos de discurso (ou seja, de conteúdo em texto ou multimídia) serão permitidos, incentivados, desincentivados ou proibidos. [...] Internamente, por sua vez, detalham procedimentos de interpretação de tais regras privadas, organizam precedentes e constroem sistemas de aplicação para dar conta de todo o conteúdo que é gerado por seus usuários. Todas essas atividades, portanto, fazem parte do referido binômio de funções: elaborar e aplicar as normas que regem e gerenciam o comportamento de usuários de determinado espaço” (Brito Cruz et al., 2023).

A divisão proposta é esquemática e na realidade encontra evidentes sobreposições e afetações. A produção se sobrepõe à arquitetura na medida em

3 O conceito de *affordances* surgiu, primeiramente, com o psicólogo estadunidense James Gibson. De acordo com o estudioso, as *affordances* não se encontram no ambiente ou nos agentes, mas surgem a partir do encontro entre eles. Um exemplo dado por Gibson é de insetos que conseguem andar sobre a água. Um lago oferecerá *affordances* de andabilidade para indivíduos dessa espécie, mas não para os demais. No mesmo sentido, *affordances* em novas tecnologias não ditam o comportamento de seus usuários, mas configuram o ambiente de uma forma que molda o envolvimento destes. Segundo Boyd (2010), quatro tipos de *affordances* surgem na arquitetura de redes sociais: (i) persistência, pois as expressões online são automaticamente registradas e arquivadas; (ii) replicabilidade, pois o conteúdo produzido pode ser duplicado; (iii) escalabilidade, pois a visibilidade potencial do conteúdo é grande; e (iv) capacidade de pesquisa, pois o conteúdo produzido por outros usuários pode ser acessado por meio de pesquisa (Nascimento et al., 2022).

que só é possível produzir quando a plataforma é desenhada com essa possibilidade, por exemplo. A moderação também incide na produção quando uma proibição gera uma autocensura em um criador de conteúdo.

A sobreposição mais complexa de organizar é aquela entre curadoria e moderação, como classificar a atividade de diminuição na recomendação de um conteúdo extremista que continue permitido, por exemplo. O corte proposto neste trabalho baseia-se na publicidade prévia dessas regras e na sua possível aplicação. Assim, a moderação de conteúdo se diferencia da curadoria em dois aspectos. Primeiro, tais regras de comportamento e adequação (em geral derivadas de preocupações de segurança e integridade do serviço e dos demais usuários) são dispositivos públicos, não se confundindo com decisões comerciais de privilegiar recomendações disto ou daquilo, ou de fomentar algum estilo específico de curadoria. Segundo, se aplicam em situações as quais a arquitetura não conseguiu prever de antemão (quando a plataforma simplesmente não disponibiliza ferramentas para postagem de vídeos, eles não estariam proibidos por regra, mas impossibilitados arquitetonicamente).

Por fim, caracterizar tais atividades não pode prescindir de uma nota sobre a escala na qual se realizam. A expansão da conectividade e da capacidade de processamento em dispositivos acessíveis a importante fração da população global, nos anos 2000 e 2010, e as oportunidades e atrativos oferecidos por tais plataformas implicaram um crescimento vertiginoso que é fato notório, tornando-

-as suas controladoras empresas de valor trilionário. Com esse crescimento, veio um volume absolutamente sem precedentes de geração de conteúdo, que fez com que tais atividades (produção, curadoria/arquitetura e moderação) ocorressem em contexto, em escala industrial e massiva⁴.

- No caso da *produção*, a escala multiplica exponencialmente as possibilidades do que será expresso por cada um dos usuários. Enorme diversidade interseccional abre imensas possibilidades de usos a partir de contextos sociais, econômicos, étnico-raciais, culturais e linguísticos – que passam a poder interagir entre si. Mesmo plataformas que oferecem serviços a determinados nichos não deixam de abarcar uma diversidade extraordinária na expressão.
- No eixo da *curadoria*, a escala é justamente o elemento que torna seus processos nevrálgicos na experiência de qualquer usuário. As enormes quantidades de spam presentes na internet são um dos elementos que podem impossibilitar que usuários encontrem aquilo que procuram, por exemplo. Essa organização ao mesmo tempo massiva e personalizada do que deve ser entregue/recomendado para cada audiência exigiu de muitos desses serviços o desenvolvimento de sistemas cada vez mais complexos de inferência sobre interesses e preferências baseados em dados pessoais.

4 Dados da edição mais recente da pesquisa Data Never Sleeps mostram, por exemplo, que a cada minuto 500 horas de vídeo são postadas no YouTube e mais de 347 mil tweets são postados. Disponível em: <https://www.domo.com/data-never-sleeps#>.

- Entendida sob este quadro, a *moderação* de conteúdo é um desafio ao mesmo tempo logístico e político/normativo. Político porque revela quais as escolhas que quem controla a plataforma faz sobre segurança e adequação, refletindo valores sociais e políticos e escolhas delicadas a respeito de questões de enorme tensão (como limites da expressão e discursos de ódio, por exemplo). Logístico porque requer que tais escolhas políticas sejam imbuídas em sistemas de detecção, avaliação e aplicação de regras que funcionem conforme quantidades colossais de denúncias, potenciais danos e fluxo de informações.

A IA E AS PLATAFORMAS: DA PRODUÇÃO À MODERAÇÃO DE CONTEÚDO

O atravessamento dos avanços tecnológicos no aprendizado de máquina ocorre em cada um dos tipos de atividades elencados acima – até porque tais processos já encontram usos correntes e atuais de ferramentas de IA. Tais usos se popularizaram conforme consolidados o desafio logístico e a dimensão industrial dessas atividades, em especial nos dois eixos que estão sob o controle direto das plataformas – o da curadoria e o da moderação.

Com efeito, sem tecnologias de aprendizagem de máquina e suas aplicações, plataformas de internet de conteúdo gerado por usuários não funcionariam da maneira como funcionam hoje. Não haveria feeds que recomendam conteúdo inferindo qual o nosso interesse ou o que

cativa a nossa atenção, não existiriam boas ferramentas de busca para encontrar vídeos ou imagens que gostaríamos de lembrar, redes sociais estariam potencialmente repletas de *spam*⁵ e conteúdo indesejado produzido por seus milhões de usuários. Não existiriam filtros ou bons mecanismos de edição de imagem e som para criadores de conteúdo, não seriam geradas legendas automáticas em vídeos e denúncias de conteúdo impróprio possivelmente demorariam muito mais para serem analisadas.

Para esquematizar o atravessamento mencionado, alguns dos usos de IA em plataformas de internet são elencados como exemplos na tabela 1.

Cada um dos eixos enseja discussões científicas relevantes no campo da regulação e de políticas públicas, em especial se consideradas as questões já mencionadas relacionadas ao tema da desordem informacional. O da produção de conteúdo, por exemplo, pode explorar tanto as possibilidades de pesquisa no âmbito da criatividade quanto a interseção com a tutela de direitos autorais, seja nas bases de treinamento ou nos resultados gerados pelas ferramentas de IA generativa. Esse eixo também pode abordar como tais conteúdos “sintéticos” serão recebidos por usuários de internet, anabolizando por

5 *Spam* é um termo em inglês utilizado para significar o envio em massa de mensagens não solicitadas pelo destinatário. Peças de spam não são, necessariamente, ilegais. Vistas individualmente, podem ser totalmente inofensivas e protegidas pela liberdade de expressão. Todavia, caso esse tipo de conteúdo passe a dominar a experiência de um usuário em uma rede social, esta poderá ter perdido sua utilidade (Monteiro et al., 2021).

TABELA 1

Usos de IA por tipo de atividade em plataformas de internet

	Produção	Curadoria/Arquitetura	Moderação
Usos de ferramentas de IA	<ul style="list-style-type: none"> • Geração de imagens e textos em IA generativa para postagem; • Edição de itens multimídia de maneira acelerada e facilitada, como em aplicativos que criam filtros de imagem, dublam vídeos ou geram legendas; • Controle de perfis, contas ou canais de forma automatizada, inclusive considerando interação com usuários humanos (como chatbots). 	<ul style="list-style-type: none"> • Seleção e organização automatizada de peças de conteúdo a serem entregues em estrutura de feed ou similar em determinada ordem, a partir de inferências feitas sobre comportamentos de usuários; • Fornecimento de perfis ou canais sugeridos ao usuário a partir de seus interesses; • Seleção e disposição de conteúdo relacionado a conteúdo consumido. 	<ul style="list-style-type: none"> • Filtragem prévia de conteúdo a ser publicado a partir de regras específicas, como o treinamento de ferramentas que previnem o compartilhamento de conteúdo que contenha material protegido por direito autoral; • Detecção de postagens com conteúdo que potencialmente viola termos de uso e ranqueamento das possibilidades de violação a partir de precedentes; • Organização de ordem de análise de conteúdos por ação humana por inferência sobre sua urgência frente a critérios de risco.

cessos de desinformação e/ou polarização política. Na frente da arquitetura e curadoria, algoritmos de recomendação sendo capazes de fazer inferências mais certas ou relevantes sob determinados propósitos.

Este artigo vai explorar de forma mais detida as possibilidades e riscos em ape-

nas um desses eixos, o da moderação de conteúdo. A escolha tem motivo na centralidade que o tema tem encontrado nas discussões regulatórias ao redor do globo, com especial eco na aprovação do Digital Services Act (DSA) pela União Europeia⁶. Normativas como o DSA exa-

6 O Digital Services Act (DSA) é um regulamento europeu cujo objetivo é regular os serviços digitais e criar um ambiente digital seguro, em que os direitos fundamentais dos usuários são respeitados. Embora não mencione explicitamente o termo “inteligência artificial”, o DSA inclui várias referências ao uso de algoritmos e sistemas automatizados. No artigo 34, por exemplo, a legislação pontua que provedores de plataforma on-line e mecanismos de busca de grande porte devem identificar, analisar e avaliar di-

ligentemente quaisquer riscos sistêmicos decorrentes da concepção, do funcionamento de seu serviço e de seus sistemas relacionados ou da utilização de seus serviços, incluindo sistemas algorítmicos. Por sua vez, os artigos 14 e 27 do DSA indicam que as plataformas são obrigadas a indicar nos termos e condições de seus serviços informações sobre moderação de conteúdo e sistemas de recomendação por algoritmos. Disponível em: <https://eur-lex.europa.eu/eli/reg/2022/2065/oj>. Acesso em: 25/jan./2024.

cerbam como a moderação de conteúdos é crescentemente percebida como uma tarefa crítica para mitigação dos danos que podem ser catalisados por ferramentas digitais.

A IA NA MODERAÇÃO DE CONTEÚDO: FUNCIONALIDADES E LIMITAÇÕES

O campo da IA soma-se nesse esforço na medida em que torna possível “ensinar” computadores a reconhecerem padrões e sinais típicos em conteúdos postados por usuários, entrando em cena quando falamos, por exemplo, das tarefas de detecção, priorização e análise de conteúdos que violam o conjunto de regras das plataformas.

Assim, é necessário destrinchar exemplificativamente a cadeia de processos que podem receber diferentes tipos de aplicações de IA. Na frente da detecção, por exemplo, é possível utilizar ferramentas de reconhecimento de padrões (em imagens ou outros conteúdos) para separá-los para uma análise. Como explica um glossário feito pela Digital Trust & Safety Partnership (DTSP), uma parceria de mais de 20 empresas donas de plataformas de internet para debaterem questões de moderação de conteúdo e segurança, a “automação pode ser usada para detectar potenciais abusos, por meio de métodos como filtragem de palavras-chave, correspondência de *hash* [que é a comparação entre o conteúdo analisado e conteúdos anteriores a partir de técnicas específicas], análise comportamental, aprendizado de máquina e inteligência artificial” (DTSP, 2023).

Em uma linha de priorização, por sua vez, é possível automatizar a análise de

critérios de prioridade na “fila” de conteúdos a serem analisados por revisores humanos, a fim de tornar seu trabalho mais eficiente ou de treinar a partir de como essa priorização pode ser otimizada. Por mais que essa discussão sobre priorização seja um dos pontos mais obscuros ao debate público, casos recentes demonstram sua importância e sensibilidade (Brito Cruz [coord.]; Lana; Jost, 2023)⁷.

Por fim, na frente da análise, máquinas podem aprender a identificar sinais de possíveis problemas a partir de bancos de dados de violações pregressas, fornecendo índices ou inferências sobre um conteúdo detectado ainda não analisado. Se a probabilidade de violação for muito alta, plataformas poderão programar seus sistemas para agir inclusive sem revisão humana prévia. Nesses casos, as definições do que é “muito alto” e os casos apropriados para esse tipo de decisão podem gerar polêmicas importantes⁸.

Por mais que sejam tarefas diferentes, tais casos se conectam: um conteúdo pode ser detectado já com um índice alto de possível violação aos termos de uso – e, conseqüentemente, priorizado nas “filas” de análise. Como explicou Emma Llansó (2020), determinadas plataformas inclusive utilizam tais índices para realizar “filtragem” e aplicação de regras no momento do upload (ou seja, no tempo entre o ato da postagem pelo usuário e a publicação

7 Como já descrevemos em trabalho sobre o caso do sistema XCheck, operado pela Meta.

8 Em caso recente, a detecção de conteúdo de nudez infantil e a tomada de decisão rápida por seu alto risco tiveram conseqüências drásticas em uma família nos EUA após uma criança cometer um erro (Hill, 2024).

do conteúdo na internet). Tais práticas surgiram com força para detectar conteúdo protegido por propriedade intelectual (Hartmann; Silva, 2020) e abuso sexual infantil, mas há pressão para o seu uso em outros contextos.

Se considerada a escala necessária para dar conta do fluxo de informações geradas nos ambientes das plataformas digitais, seria impossível contar com processos “artesanais”, ou seja, sem qualquer tipo de aprendizado de máquina. Os serviços e as qualidades por eles oferecidos estariam possivelmente inviabilizados e os sistemas de segurança seriam significativamente mais custosos. Em outro trabalho desenvolvido no âmbito do InternetLab, explicamos como “entre as boas práticas do setor está a composição equilibrada e complementar entre sistemas automatizados (e ‘inteligentes’) e a participação de colaboradores humanos que supervisionam, revisam e direcionam o trabalho de máquinas” (Brito Cruz et al., 2023).

RISCOS E SUAS ESTRATÉGIAS DE MITIGAÇÃO

Abuso, erros e problemas ocorrem na moderação de conteúdo – e nem sempre eles estão ligados ao uso da IA⁹. Mesmo que uma miríade de casos ocorra por má prática corporativa e ética, falta de investimento e erro humano, é certo que também toda ferramenta de IA poderá vir a cometer erros e apresentar vieses inde-

sejados¹⁰. Pesquisas continuamente têm demonstrado como tais tecnologias podem reproduzir desigualdades estruturantes e potencializar violências preexistentes. Sua natureza é de serem produzidas e terem seu treinamento desenhado por pessoas que carregam em si mesmas experiências e perspectivas.

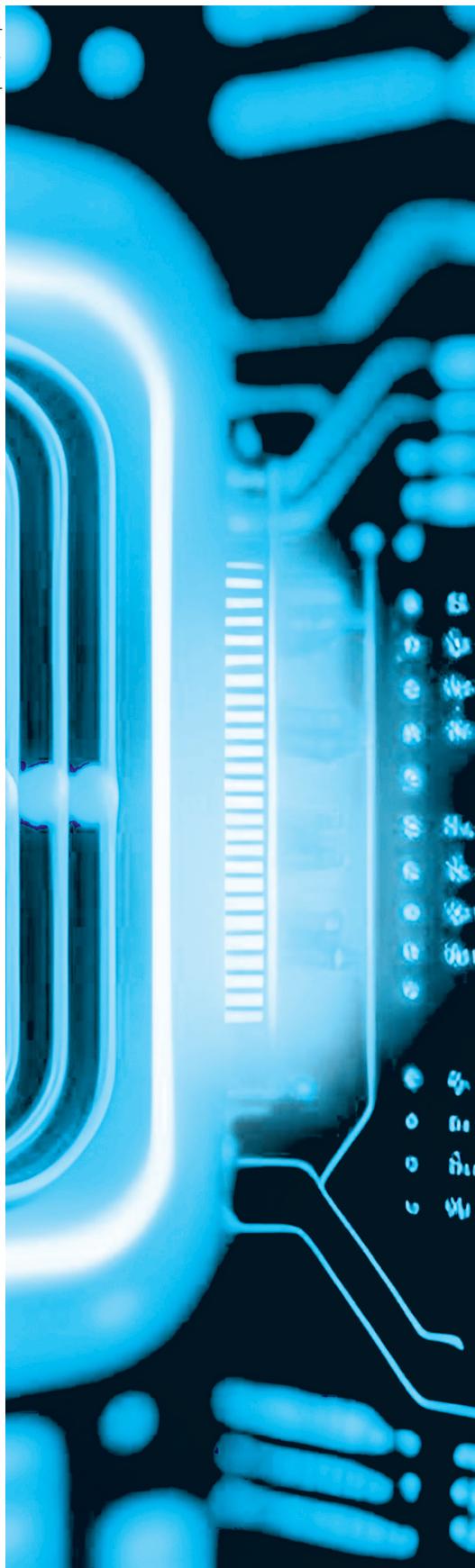
Com efeito, um dos problemas está relacionado às bases de dados dedicadas aos treinamentos de aprendizado de máquina. Tais bases podem possuir uma série de limitações ou mais dados sobre um fenômeno e não outro, bem como podem não estar atualizadas. Se uma máquina aprende padrões a partir do comportamento de usuários de um tipo de serviço, por exemplo, isso pode fazer com que não esteja preparada para lidar com uma demografia de usuários que usam tal serviço de maneira diferente, por qualquer motivo que seja.

Algumas evidências da detecção de vieses em sistemas de moderação de conteúdo que usam IA são relevantes para servir de alerta e projetar formas de mitigação a serem adotadas.

- O primeiro foi detectado por pesquisa realizada pelo InternetLab (Dias Oliva; Antonialli; Gomes, 2021; 2019). Neste caso, pesquisadores testaram uma ferramenta criada por uma subsidiária do Google para análise do “nível de toxicidade” de discursos em textos. O teste comparou como a tecnologia classificava escritos em inglês postados por *drag*

9 Como organizado em bom mapeamento em: Silva; Cesar (2022).

10 Para mais informações sobre viés algorítmico, ver: Noble (2018).



queens, de um lado, e por reconhecidos nacionalistas brancos dos EUA, de outro. O achado é de que esse tipo de tecnologia pode não ser capaz de distinguir nuances de contextos, como, por exemplo, a diferença entre discursos de ódio contra pessoas LGBTQIA+ e conteúdo publicado por pessoas LGBTQIA+, que frequentemente ressignificam termos compreendidos como ofensivos, positivando-os para comunicação entre pares. Os perfis de *drag queens* foram considerados potencialmente mais tóxicos do que os perfis de supremacistas brancos, pois termos comumente usados por pessoas LGBTQIA+, como *sissy*, *gay*, *lesbian* e *queer* eram considerados potencialmente tóxicos pela ferramenta de IA, independentemente do contexto. O caso da pesquisa sobre *drag queens* demonstra distorções e vieses da IA, que não foi capaz de identificar as matizes específicas da comunicação de pessoas *queers*, em que o uso de uma linguagem “pseudo-ofensiva” é uma forma comum de comunicação entre membros da própria comunidade. Pesquisadores como Udupa, Maronikolakis e Wisiorek (2023) argumentam sobre a dificuldade de classificar e treinar sistemas para reconhecer “discurso extremo” mesmo dentro de um mesmo idioma.

- Um segundo exemplo se refere a detecção e moderação da nudez em plataformas de internet. Algumas plataformas proíbem, de forma legítima, a disseminação de imagens contendo partes de um corpo desnudo, como, por exemplo, a Meta. Isso ocorre não só por decisões de negócio sobre qual o tipo de conteúdo erótico é ou não aceitável, mas também porque uma

escolha como essa pode reduzir riscos de circulação de conteúdos de exploração sexual de crianças e adolescentes pela dificuldade de aferição de idade precisa em imagens e vídeos¹¹. Ao mesmo tempo, quando há previsão do banimento desse tipo de conteúdo, empresas podem elencar exceções a essas políticas de imagens com nudez que façam referência a (i) momentos de parto e pós-parto, (ii) procedimentos cirúrgicos e médicos de confirmação de gênero, e (iii) autoexames de câncer ou publicações sobre prevenção e avaliação de doenças em genitálias e partes íntimas¹². No entanto, existem casos recorrentes de remoção de conteúdo, especialmente de pessoas trans e não binárias, com publicações sobre procedimentos de afirmação de gênero, com base nas políticas de nudez e de proposta de cunho sexual. Em resposta às remoções, pessoas trans e não binárias engajaram-se no movimento #DeserveToBeHere (em tradução livre, #MerecemosEstarAqui), em que contestavam o banimento de fotos com nudez sobre transição de gênero¹³.

11 Em resposta ao Comitê de Supervisão do Facebook, a Meta afirmou, por exemplo, que na elaboração dos princípios gerais de suas políticas sobre nudez considerou: "(1) a natureza privada ou sensível das imagens; (2) se foi dado consentimento para a obtenção e compartilhamento de imagens de nudez; (3) o risco de exploração sexual; e (4) se a divulgação das imagens pode levar a assédio fora da plataforma, particularmente em países onde elas podem ser culturalmente ofensivas". Disponível em: <https://oversightboard.com/decision/BUN-IH313ZHJ/>.

12 Política de nudez adulta e atividades sexuais da Meta, disponível em: <https://transparency.fb.com/en-gb/policies/community-standards/adult-nudity-sexual-activity/>.

13 Disponível em: <https://www.thepinknews.com/2021/04/22/instagram-trans-bodies-censorship-we-deserve-to-be-here/>.

- Um terceiro exemplo é dado pelo desafio colocado a partir da diversidade de idiomas a serem analisados na moderação de conteúdo. Recentes relatórios (Duarte; Llansó; Loup, 2017; Nicholas; Bhatia, 2023) do *think tank* Center for Democracy and Technology demonstram como tecnologias de reconhecimento e análise de linguagem (como grandes modelos de linguagem que chamam a atenção a partir de 2022) são desproporcionalmente testados em inglês, o que limita seu uso no processamento de conteúdos – e possivelmente ajuda na moderação.

Riscos como esse aquecem o debate sobre quais mecanismos de mitigação – ou segurança – devem ser cobrados como medidas protetivas razoáveis para os usuários e seus direitos. Vistas sob diversos prismas, tais questões levantam tanto propostas de banimento de determinadas possibilidades (inclusive no âmbito da própria arquitetura dos serviços das plataformas, não só de sua moderação) como visões mais *laissez-faire*, que confiam no aperfeiçoamento contínuo das máquinas.

Três vertentes mais pragmáticas se destacam. Não esgotam as possibilidades e podem se combinar, organizando caminhos não excludentes para enfrentar o problema.

Em primeiro lugar há a defesa por maior transparência na moderação de conteúdo, com ampliação de formas de prestação de contas e da consolidação de direitos de usuários por um “devido processo” na análise de suas atividades. Buscando aumentar o acesso à informação e produzir uma moderação com maior legitimidade e justificação, essa vertente contempla proposições de normas e ter-

mos de uso sobre o que é proibido e sobre como será a análise de eventual violação, escritos em linguagem clara e acessível, traduzidos para os idiomas de todas as regiões onde são navegados e dispostos de maneira que qualquer pessoa possa encontrá-los, por exemplo. Propõe também que existam métodos de prestação de contas nos quais pessoas que acessam as redes sociais possam dar opiniões e feedbacks sobre o exercício da moderação. Contempla também a ideia de que as plataformas possuam instâncias recursais, de maneira a possibilitar que as pessoas possam, por meio de procedimentos claros e formais, conhecer a razão das decisões que influenciam sua expressão nas redes, e contestá-las, caso julguem justo e necessário. Algumas dessas propostas inclusive visam submeter tais recursos à revisão humana.

Na esteira desse último componente, a segunda vertente é sobre a incorporação de etapas de supervisão realizadas por pessoas qualificadas que revisem e direcionem o trabalho das máquinas nas mais variadas etapas. A inserção desse tipo de camada pressupõe que esses processos precisam ser complementados com um elemento humano que os aprimore e agregue nuances, freios e contrapesos às suas análises para abrandar eventuais vieses. Tal revisão adicionaria matizes às decisões maquinadas e somaria especificidades e contextos que podem ser peças-chave para uma gestão saudável do discurso que circula nas plataformas ou para a proteção de populações historicamente marginalizadas.

Uma terceira vertente pragmática é a que contempla propostas para o constante investimento, reavaliação e inovação tec-

nológica. Com efeito, sistemas automatizados para moderação de conteúdo não só precisam de constante mirada por serem parte do produto (e valor) das plataformas, mas também porque as ameaças não são estáticas. Frente à enorme quantidade de conteúdo, de sua velocidade de circulação e das frequentes mudanças de contextos que envolvem a expressão, é necessário que as plataformas dediquem um olhar cuidadoso e continuado para essa questão. Isso deve envolver, por exemplo, diálogo com especialistas locais, equipes interdisciplinares com pontos focais regionais atentos e especializados, capazes de abordar desafios atrelados a idiomas e conjunturas específicos, como um processo eleitoral, a promulgação de determinada legislação ou o início de um conflito. Neste viés, percebe-se especial atenção à existência de métricas específicas e transparentes de análise, e que todos esses mecanismos combinados sejam capazes de fomentar melhorias nas ferramentas tecnológicas e, por consequência, na gestão da moderação de conteúdo.

Um dos exemplos marcantes nessa última vertente vem da sugestão dada por pesquisadores da OpenAI sobre o uso do seu GPT-4 na moderação de conteúdo (Weng; Goel; Vallone, 2023). Na apresentação dessa alternativa, os seus defensores explicam que conseguiram reduzir significativamente o tempo de “curva de aprendizado” dos sistemas de moderação na implementação de novas regras por conta do uso do modelo, ainda que ele conseguia detectar pontos de aperfeiçoamento nas regras para sua aplicação consistente. Em uma coleta de opiniões de *experts* sobre o experimento, o colunista Casey

Newton (2023) apontou que a promessa é bem vista, mas gera uma série de questionamentos: se, por um lado, recursos humanos podem ser mais bem aplicados em tarefas mais complexas e específicas, por outro, o uso de IA generativa para moderação de conteúdo pode inclusive criar problemas de “explicabilidade” nas decisões por remoção, o que pode gerar problemas legais para a plataforma perante a nova lei europeia, o DSA.

CONSIDERAÇÕES FINAIS

A promissora evolução das tecnologias de IA deve acentuar ainda mais as tendências mapeadas neste modesto artigo em todos os eixos aqui elencados. As possibilidades para produção de conteúdos por ferramentas de IA generativa, por exemplo, já são um importante ponto de atenção para reguladores ao redor do mundo, inclusive na seara eleitoral. Inovação atrás da outra, crescem os riscos de conteúdos “sintéticos” se passarem por “orgânicos” (ou seja, não tratados ou gerados por mecanismos de IA) e engatilharem diferentes crises. Esse tema está no

centro de incipientes – porém aceleradas – discussões regulatórias (em especial de autoridades eleitorais, como o Tribunal Superior Eleitoral brasileiro), com alternativas que vão do banimento à rotulagem obrigatória de conteúdos.

No plano da curadoria/arquitetura, tais grandes modelos de linguagem cada vez mais estão embarcados na experiência dos usuários, potencialmente reestruturando sua dieta de informações. Na moderação, como vimos, o uso de ferramentas de ponta promete, mas requer supervisão, transparência e investimento para ser mais do que uma maneira de diminuir custos de uma atividade bastante sensível a direitos humanos.

A incerteza gerada nesse campo não é necessariamente mais impressionante do que aquela produzida no mundo do trabalho, dos negócios ou das mudanças climáticas, mas tem uma característica especial. Considerando a forma como a população mundial hoje conversa, se informa e se polariza, ela se esparrama sobre todas as outras. As capacidades de criarmos ambientes para expressão e troca, para informar e ser informado, é que estão em jogo.

REFERÊNCIAS

- BRITO CRUZ, F. (coord.), LANA, A. de P.; JOST, I. *Iguais perante as plataformas? Equidade e transparência na moderação de conteúdo em plataformas digitais*. São Paulo, InternetLab, 2023. Disponível em: https://internetlab.org.br/wp-content/uploads/2023/08/relatorio_internetlab_crosscheck_PORTUGUES_ok2.pdf.
- BRITO CRUZ, F. et al. Contribuição do InternetLab ao Expediente T-8.764.298 Acción de Tutela instaurada por Esperanza Gómez Silva contra Facebook Colombia S.A.S, Instagram Colombia y Meta Platforms, Inc., Ministerio de las Tecnologías de la Información y las Comunicaciones (MinTIC) y Superintendencia de Industria y Comercio. São Paulo, InternetLab, 2023.
- BRITO CRUZ, F. et al. *Internet e eleições no Brasil: diagnósticos e recomendações*. São Paulo, InternetLab, 2019. Disponível em: http://www.internetlab.org.br/wp-content/uploads/2019/09/policy-infopol-26919_4.pdf.
- DIAS OLIVA, T.; ANTONIALI, D. M.; GOMES, A. "Fighting hate speech, silencing drag queens? Artificial intelligence in content moderation and risks to LGBTQ voices online". *Sexuality & Culture*, v. 25, 2021, pp. 700-32.
- DIAS OLIVA, T.; ANTONIALI, D. M.; GOMES, A. "Drag queens e inteligência artificial: computadores devem decidir o que é 'tóxico' na internet?". InternetLab, 28/jun./2019. Disponível em: <https://internetlab.org.br/pt/noticias/drag-queens-e-inteligencia-artificial-computadores-devem-decidir-o-que-e-toxico-na-internet/>
- DTSP – Digital Trust & Safety Partnership. *Trust and safety glossary of terms* Jul./2023. Disponível em: https://dtspartnership.org/wp-content/uploads/2023/07/DTSP_Trust-Safety-Glossary_July-2023.pdf.
- DUARTE, N.; LLANSÓ, E.; LOUP, A. "Mixed messages? The limits of automated social media content analysis". Center for Democracy & Technology, 2017. Disponível em: <https://cdt.org/wp-content/uploads/2017/11/2017-11-13-Mixed-Messages-Paper.pdf>.
- HARTMANN, I. A.; SILVA, L. A. da. "Inteligência artificial e moderação de conteúdo: o sistema CONTENT ID e a proteção dos direitos autorais na plataforma YouTube". *IUS Gentium*, v. 10, n. 3, 2020, pp. 145-65.
- HILL, K. "Como um erro on-line de seu filho pode arruinar a vida digital dos pais". *O Estado de S. Paulo*, 2024. Disponível em: <https://www.estadao.com.br/lifestyle/como-um-erro-on-line-de-seu-filho-pode-arruinar-a-vida-digital-dos-pais>.
- LLANSÓ, E. "No amount of 'AI' in content moderation will solve filtering's prior-restraint problem". *Big Data & Society*, v. 7, n. 1, 2020.
- MONTEIRO, A. et al. *Armadilhas e caminhos na regulação da moderação de conteúdo. Diagnósticos & recomendações*. São Paulo, InternetLab, 2021. Disponível em: https://internetlab.org.br/wp-content/uploads/2021/09/internetlab_armadilhas-caminho-moderacao.pdf.
- NASCIMENTO, L. et al. "Públicos refratados: a atuação de grupos de extrema direita brasileiros na plataforma Telegram". *Internet & Sociedade*, v. 3, n. 1, 2022.
- NICHOLAS, G.; BHATIA, A. "Lost in translation: large language models in non-english content analysis". *Center for Democracy & Technology*, 2023. Disponível em: <https://cdt.org/insights/lost-in-translation-large-language-models-in-non-english-content-analysis/>.

- NEWTON, C. "OpenAI wants to moderate your content". *Platformer*, 2023. Disponível em: <https://platformer.substack.com/p/openai-wants-to-moderate-your-content>.
- NOBLE, S. U. *Algorithms of oppression: How search engines reinforce racism*. Nova York, New York University Press, 2018.
- SILVA, S. P. da; CESAR, D. J. T. "Inteligência artificial, moderação de conteúdos no YouTube e a proteção de direitos: características, problemas e impactos políticos". *Liinc em Revista*, v. 18, n. 2, nov./2022, pp. 1-21.
- UDUPA, S.; MARONIKOLAKIS, A.; WISIOREK, A. "Ethical scaling for content moderation: extreme speech and the (in)significance of artificial intelligence". *Big Data & Society*, v. 10, n. 1, 2023.
- WARDLE, C.; DERAKSHAN, H. "Information disorder: toward an interdisciplinary framework for research and policy making". *Council of Europe*, 27/set./2017. Disponível em: <https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-research/168076277c>.
- WENG, L.; GOEL, V.; VALLONE, A. "Using GPT-4 for content moderation". *OpenAI Blog*, 2023. Disponível em: <https://openai.com/blog/using-gpt-4-for-content-moderation>.

Starline/freepik

Inteligência artificial, urbanização e cidades

Márcia Regina Martins Martinez Corso

Luísa Paseto

André Carlos Ponce de Leon Ferreira de Carvalho

resumo

A forte onda de urbanização criou uma demanda mundial pela melhoria da qualidade de vida nas cidades, provocando grandes avanços nas áreas de planejamento urbano, geografia social e tecnologias da informação e comunicação. A inteligência artificial vem desempenhando um papel crucial na redefinição do planejamento urbano e na criação de cidades mais inteligentes, sustentáveis e adaptadas às necessidades de seus habitantes. Os benefícios de uma inteligência artificial ética e responsável já são observados nas cidades com o desenvolvimento de infraestruturas urbanas e tecnológicas sustentáveis, proporcionando melhor qualidade de vida para a população.

Palavras-chave: cidades inteligentes; urbanização; cidades; cidades sustentáveis; planejamento urbano.

abstract

The strong urbanization wave created a worldwide demand for improving quality of life in cities, resulting in major advances in the areas of urban planning, social geography and information and communication technologies. Artificial intelligence is playing a crucial role in redefining urban planning and creating cities that are smarter, more sustainable cities and adapted to the needs of their inhabitants. The benefits of ethical and responsible artificial intelligence are already observed in cities with the development of urban and technological sustainable infrastructure, providing a better quality of life for the population.

Keywords: smart cities; urbanization; cities; sustainable cities; urban planning.



o impacto da inteligência artificial na vida das pessoas cresce em intensidade e abrangência. Embora isso ocorra em várias áreas de aplicação, o Brasil, em sua chamada para centros nacionais de pesquisa aplicada em IA, como parte da Estratégia Brasileira de IA, priorizou as áreas de agronegócios, cidades inteligentes,

indústrias, saúde e segurança cibernética.

A aglutinação de pessoas em cidades é consequência direta da necessidade de sobrevivência. As primeiras cidades foram formadas por volta de 6.000 a.C. na região da Mesopotâmia, que hoje inclui o Iraque e regiões de países próximos no oeste da Ásia. Embora não exista um consenso acerca de qual foi a primeira cidade, uma das principais candidatas é a cidade de Uruk, no

atual Iraque, que se estima ter tido cerca de 60 mil habitantes (Metropolitan, 2003).

Como consequência da expansão territorial de aglomerados urbanos, do aumento da população mundial e de mudanças nas relações sociais, mais pessoas moram em cidades. Transformações sociais e tecnológicas que acompanharam a evolução da humanidade resolveram e reduziram vários problemas das

MÁRCIA REGINA MARTINS MARTINEZ CORSO

é pesquisadora do Centro de Inteligência Artificial Aplicada em Cidades Inteligentes (Iara), apoiado pela Fapesp, MCTI e CGI.br.

LUÍSA PASETO é coordenadora da plataforma *inteli.gente* para cidades inteligentes e de análise de indicadores de cidades para o centro Iara.

ANDRÉ CARLOS PONCE DE LEON FERREIRA

DE CARVALHO é professor titular e diretor do Instituto de Ciências Matemáticas e de Computação, da USP São Carlos, e coordenador do centro Iara.

idades, mas trouxeram novos problemas que põem em risco o futuro do planeta.

Nas últimas décadas, a IA tem ajudado a melhor entender vários problemas, suas causas, dinâmicas e efeitos, além de permitido o desenvolvimento de ferramentas computacionais que resolvem ou minimizam vários deles, com benefícios em áreas como educação, mobilidade, saúde, segurança e sustentabilidade. Por isso, ferramentas de IA são importantes aliadas para lidar com os problemas das nossas cidades e ajudar a evitar ou reduzir os riscos dessas questões, com potencial para tornar as cidades mais eficientes, inteligentes, sustentáveis e centradas nas necessidades dos cidadãos. Este texto discorre sobre como a IA pode tornar melhor as nossas cidades e como avaliar a situação das cidades brasileiras no que se espera de uma cidade inteligente e sustentável.

CIDADES

Mas o que é uma cidade? De acordo com o UN-Habitat (2020), existem várias definições diferentes, que variam entre países e regiões, tornando difícil oferecer uma definição única, que seja universalmente aplicável. As definições existentes variam entre aquelas que utilizam um único critério e aquelas que utilizam vários critérios, que podem inclusive ser conflitantes. Neste estudo, é reportado que, de acordo com uma pesquisa com os 233 países que fazem parte da ONU, 104 adotam apenas um critério, que pode ser a função administrativa, o tamanho ou a densidade da população ou as características urbanas. Nele, é mencionado que o Brasil define cidades como “centros administrativos de municípios e distritos, incluindo zonas rurais”.

Para atender ao Objetivo de Desenvolvimento Sustentável (ODS) 11, Cidades e Comunidades Sustentáveis¹, é necessário que seja adotada uma definição universal do que vem a ser uma cidade, e que deve ser baseada em indicadores globais. Após uma análise dos vários conceitos adotados e de uma consulta a diversos especialistas, o estudo propõe a adoção de duas definições. A primeira, sugerida por pesquisadores da Universidade de Nova York, nos Estados Unidos, tem dois componentes: a extensão urbana (área total construída contígua) e o espaço aberto urbanizado (área total não construída dentro de áreas construídas ou dentro de sua vizinhança imediata) (Angel, 2016). O espaço aberto urbanizado inclui áreas livres, florestas e parques. Os limites das cidades são identificados por meio da análise de imagens obtidas por satélites. A segunda definição, da União Europeia, é baseada em uma variação do grau de urbanização (Degurba, do original em inglês *degree of urbanization*) (Dijkstra, 2014). Ela utiliza como unidade de análise o nível administrativo local 2 (LAU2, do original em inglês *local administrative unit level 2*), que distingue um assentamento em três categorias, de acordo com a contiguidade dos assentamentos e do tamanho e densidade da população: densa, intermediária e leve, com grades populacionais de 1 km². Com essas categorias, um assentamento pode ser classificado em centro urbano (ou grupo de alta densidade), grupo urbano ou célula de grade rural. Independente da

1 Ver: <https://brasil.un.org/pt-br/sdgs>.

2 Ver: <https://unece.org/housing/smart-sustainable-cities>.

definição adotada, cidades são locais em que pessoas moram e/ou trabalham e têm acesso a serviços públicos.

Após a Segunda Guerra Mundial, com a destruição de várias cidades e as grandes ondas migratórias, ocorreram grandes transformações socioeconômicas que reforçaram a necessidade de reconstrução e de transformação das cidades. Transformações socioeconômicas também levaram a uma aceleração do crescimento do número e do tamanho das cidades, alterando drasticamente a taxa de urbanização do planeta.

URBANIZAÇÃO

Por volta do ano de 2008, a quantidade de pessoas vivendo em cidades, a população urbana, ultrapassou a população rural (Townsend, 2013). De acordo com a ONU (United Nations, 2019), que apresenta uma análise do crescimento e um prospecto do processo de urbanização no mundo até 2050, em 2018, aproximadamente 55% da população mundial estava em assentamentos urbanos. Essa presença se distribui de uma forma muito desigual no tamanho das cidades; praticamente metade da população mundial vive em cidades com mais de 500 mil habitantes.

A renda *per capita* influenciou o período de urbanização dos países, que começou mais rapidamente nos de renda mais elevada e deve ser mais rápido agora nos países de renda mais baixa. A urbanização é maior na América do Norte e na América Latina, em torno de 80% da população, com a Europa vindo em seguida, com 74%. A urbanização é menor na África abaixo da região do Saara, 38%, e na Ásia, 48%, vindo

logo depois o Norte da África, com 52%. Além disso, 90% da população rural vive na África e na Ásia, regiões que, até 2050, deverão responder por 90% do crescimento da população urbana mundial.

O mesmo estudo prevê que, em 2030, 60% da população viverá nas cidades e, em 2050, essa taxa será de 68%, quando o mundo deverá ter 43 megacidades, com mais de 10 milhões de habitantes, a maioria delas nas regiões em desenvolvimento. Cerca de 80% do crescimento da urbanização até 2050 ocorrerá nas regiões de menor renda e, infelizmente, em áreas mais sujeitas a desastres ambientais e mais vulneráveis às mudanças climáticas, aumentando a desigualdade social.

CIDADES INTELIGENTES

O aumento da urbanização criou uma forte demanda pela melhoria da qualidade de vida nas cidades, que levou a grandes avanços nas áreas de arquitetura, planejamento urbano, geografia social e tecnologias construtivas (Nath, 2023). Com os avanços tecnológicos ocorridos nas últimas décadas, em praticamente todas as áreas do conhecimento, as cidades passaram a dispor de moradias mais baratas, seguras e salubres, na esteira de uma maior oferta de empregos, com um número crescente de funções e um real crescimento do alcance, da qualidade e da quantidade de serviços públicos demandados e oferecidos.

No entanto, mesmo com esses avanços, a qualidade de vida nas cidades ainda é ruim. Com a urbanização desordenada, os municípios não conseguem atender às necessidades básicas da população, causando, ou

umentando, problemas sociais e ambientais, como o desemprego, a criminalidade, a favelização e a poluição do ar e da água.

Um melhor uso das tecnologias é necessário, e isso é possível com a evolução observada nas tecnologias de informação e comunicação (TICs), em especial na inteligência artificial. Essa evolução permitiu o projeto e a implementação de novas funcionalidades, que estão por trás das cidades inteligentes e sustentáveis.

De acordo com a definição validada pelos parceiros da Unidos para Cidades Inteligentes e Sustentáveis (U4SSC – United for Smart Sustainable Cities)², iniciativa associada à ONU, uma cidade inteligente e sustentável é uma cidade inovadora que utiliza as TICs e outros meios para melhorar a qualidade de vida, a eficiência da operação e dos serviços urbanos e a competitividade, garantindo esses itens ao mesmo tempo que satisfaz às necessidades das gerações presentes e futuras nos aspectos econômicos, sociais, ambientais e culturais.

Uma das primeiras experiências de cidades inteligentes ocorreu na cidade de Los Angeles, na Califórnia, Estados Unidos, em um projeto que teve início no final da década de 1960. Nessa época, o Escritório de Análise de Comunidade (Cabla, 1974) começou a utilizar bancos de dados (uma das principais subáreas da computação), análise de agrupamentos (uma das principais subáreas da estatística e da computação, em particular, da IA) e fotografias aéreas em infravermelho para extrair, armazenar e gerenciar dados e gerar relatórios sobre a demografia dos bairros e a qualidade das habitações. O principal objetivo desse projeto era melhor direcionar recursos públicos para evitar pragas e combater a pobreza na cidade.

O que são cidades inteligentes e sustentáveis para o Brasil?

No Brasil, cidades inteligentes são as que se comprometem com um desenvolvimento urbano sustentável e abraçam a transformação digital em todas as suas dimensões econômicas, ambientais e socioculturais. Destacam-se por agir de forma planejada, inovadora, inclusiva e conectada em rede. Buscam promover o letramento digital, a governança colaborativa e uma gestão participativa. Utilizam tecnologias para resolver problemas reais, criando oportunidades e oferecendo serviços de maneira eficiente. Tudo isso com o propósito de reduzir desigualdades, aumentar a resiliência e melhorar a qualidade de vida de todos os cidadãos. Um dos pilares dessas cidades é garantir o uso seguro e responsável dos dados e das tecnologias da informação e comunicação, visando ao bem-estar coletivo (Carta, 2021).

Para que uma cidade seja considerada inteligente, alguns pilares fundamentais devem ser contemplados:

- tecnologia: a infraestrutura digital e a conectividade, essenciais para coletar e analisar dados em tempo real, possibilitando uma gestão eficiente dos recursos públicos;
- sustentabilidade: a busca por soluções ecologicamente corretas, crucial para reduzir a pegada ambiental das cidades, garantindo um futuro resiliente para as próximas gerações;
- participação cidadã: o envolvimento da população nas decisões e processos é essencial para criar uma cidade inclusiva e atender às necessidades de todos;

- eficiência: a otimização dos recursos financeiros, energéticos ou de transporte é uma marca das cidades inteligentes e sustentáveis;
- qualidade de vida: melhoria do bem-estar dos cidadãos, proporcionando uma experiência urbana mais agradável e segura.

As cidades inteligentes e sustentáveis são aquelas que utilizam tecnologias avançadas para melhorar a qualidade de vida das pessoas e tornar a gestão municipal mais eficiente e eficaz. A inteligência artificial é uma dessas tecnologias, capaz de transformar as cidades em lugares mais inteligentes e conectados.

Ela tem um grande potencial para melhorar a qualidade de vida em cidades inteligentes, otimizando a mobilidade urbana, a segurança pública, a saúde pública, a gestão de resíduos sólidos, entre outros. No entanto, é importante que a utilização seja feita de forma ética e transparente. Os municípios que implementam recursos inovadores e sustentáveis para melhorar a qualidade de vida dos cidadãos se destacam por suas iniciativas em áreas como infraestrutura inteligente, governança eficiente, mobilidade urbana e uso de tecnologias da informação para aprimorar os serviços públicos. No Brasil, análises extraídas da plataforma *inteli.gente*, indicam que um gargalo a ser desenvolvido é o planejamento da cidade, sendo comum as cidades brasileiras executarem uma ação e depois se ocuparem com legislações, regularizações, entre outras ações necessárias. Também o crescente investimento em redes de internet das coisas (IoT) vem tornando as cidades cada vez mais digitalizadas, o que permite a extração de informações customizadas, que precisam ser tratadas para a

construção de um sistema de governança e gestão que torne a cidade inteligente e provedora de qualidade de vida para a população.

REDE IARA

Como parte da Estratégia Brasileira de IA, o Ministério da Ciência, Tecnologia e Inovação (MCTI), a Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp) e o Comitê Gestor da Internet (CGI.br) lançaram em 2021 uma chamada pública nacional para a seleção dos primeiros centros de pesquisa aplicada em inteligência artificial (CPA-IA) a serem apoiados pelo governo federal. As áreas dos centros deveriam ser agronegócios, cidades inteligentes, indústrias ou saúde. Na área de cidades inteligentes foi selecionada a rede Inteligência Artificial Recriando Ambientes (Iara). A rede Iara foi proposta com o apoio de cerca de 40 instituições de ciência e tecnologia (ICTs) nacionais, distribuídas em 21 estados nas cinco regiões do país, com o apoio de empresas, cidades brasileiras e ICTs do exterior.

O primeiro projeto de cidades inteligentes e sustentáveis da rede Iara foi a cidade de Canaã dos Carajás, no Pará, sede da maior mina de ferro do mundo. A prefeitura de Canaã dos Carajás, preocupada com o futuro da cidade após o final da extração de minério de seu subsolo, havia decidido utilizar parte dos *royalties* recebidos da mineração em um fundo de desenvolvimento sustentável, que deveria ser utilizado para diversificar a economia do município e melhorar a qualidade de vida de sua população.

No projeto com Canaã, a rede desenvolveu ferramentas baseadas em IA para aproximar o cidadão da gestão municipal e para abordar

problemas do município nas áreas de educação, saúde, segurança e meio ambiente.

Outro projeto está sendo executado na cidade de Guarapuava, no Paraná, na área de saúde para a prevenção e o diagnóstico precoce de doenças raras e complexas. Para isso, está sendo implantado um laboratório capaz de monitorar por 15 anos a saúde de 4.500 pessoas, coletando dados clínicos, de dieta e de atividades físicas.

Além desses, em outros projetos, a rede Iara colabora com o MCTI na coleta e análise de indicadores de desempenho, capazes de diagnosticar o nível de maturidade das cidades brasileiras para transformação digital e desenvolvimento urbano sustentável, contribuindo com estimativas e previsões em relação ao que se espera, no Brasil, de uma cidade inteligente sustentável (inteli.gente, 2023).

PLATAFORMA INTELI.GENTE

A partir de 1950, o inchaço urbano, por conta da migração da população para a periferia das cidades em busca de oportunidades apresentadas pelo processo de industrialização, vem provocando uma concentração desigual de recursos e serviços e, em alguns casos, o crescimento sem um planejamento eficaz, o que leva a problemas de mobilidade urbana, ambientais e socioeconômicos (moradia e emprego).

O surgimento e o crescimento da periferia urbana provocam o aumento do número de pessoas atuando no mercado de trabalho informal, devido à falta de oportunidades no mercado de trabalho e à baixa qualificação profissional.

Como nas diferentes regiões do mundo, no Brasil a urbanização também ocorreu de forma desigual. Segundo o IBGE (2023), a

região mais urbanizada do Brasil é o Sudeste, onde 93% dos habitantes vivem nas cidades. O Nordeste, embora concentre o maior número relativo de municípios, possui 73% de sua população nas cidades, sendo a região menos urbanizada do país.

Para orientar as cidades brasileiras em gestão e governança para a transformação digital e o desenvolvimento sustentável, foi criada em 2019, por demanda do MCTI, a plataforma *inteli.gente*³, uma ferramenta para gestão e governança, on-time, on-line, para que os 5.570 municípios brasileiros pudessem obter diagnósticos e recomendações para transformação digital e desenvolvimento sustentável.

Indicadores de desempenho metrificados, aderência nas normas nacionais e internacionais e a Carta Brasileira para Cidades Inteligentes, entre outros, formam o conjunto de critérios para obtenção do diagnóstico e determinação do nível de maturidade nesse sistema de gestão e governança. Por meio de quatro dimensões de análises (econômica, ambiental, sociocultural e de capacidades institucionais), divididas em 31 temas de políticas públicas, um conjunto de indicadores disponibiliza diagnósticos e recomendações customizados para cada cidade (MCTI, 2023). Em 2023, a plataforma *inteli.gente* tornou-se caso de uso brasileiro do Suplemento Y 78 da ITU (2023), ao identificar e diagnosticar desafios e oportunidades de cidades inteligentes sustentáveis, com base na série de Recomendações ITU-T Y.4900:04:06 e ainda aderentes às ISOs 37120 (ISO 37120, 3018), 37122 (ISO 37122, 3019), 37123 (ISO 37123) e aos 17 ODS.

3 Ver: <https://inteligente.mcti.gov.br>.

CONSIDERAÇÕES FINAIS

Estudos mostram a degradação das condições de vida no planeta, o que repercutirá na qualidade de vida nas cidades. Ações em prol da sustentabilidade são essenciais para garantir uma boa qualidade de vida à população e sua resiliência para as gerações futuras. Os sistemas baseados em IA podem ser utilizados para monitorar indicadores associados à sustentabilidade e qualidade de vida nas cidades, predizendo ocorrências e reduzindo a incidência e o efeito de desastres sociais, econômicos e ambientais.

É importante observar que uma cidade inteligente quebra paradigmas de um Estado provedor e acentua a participação pública com inclusão social e digital, voltada para a melhoria da qualidade de vida em ambientes urbanos e rurais. Porém, o uso da IA em cidades inteligentes e sustentáveis também

apresenta riscos, que precisam ser evitados ou mitigados por meio de uma regulação da IA. Esses riscos incluem sistemas enviesados ou preconceituosos, que podem restringir o acesso da população a serviços de saúde e levar a sistemas de monitoramento de segurança tendenciosos, e o uso de um sistema de pontuação dos cidadãos, que pode puni-los ou limitar seu acesso a serviços públicos.

Ainda, a crescente urbanização traz consigo problemas para a oferta de serviços públicos de qualidade e, como efeito, o esvaziamento de ações em pequenas cidades, com perdas não somente para a qualidade de vida da população, bem como para a transformação digital e o desenvolvimento sustentável. O Brasil, especificamente em infraestrutura urbana e de conectividade para a transformação digital e o desenvolvimento sustentável, apresenta ainda muitos desafios e oportunidades a serem superados e aproveitados para se tornar inteligente e não apenas digital.

REFERÊNCIAS

- ANGEL, S. et al. *Atlas of urban expansion*. Cambridge, Ed. Lincoln Institute of Land Policy, 2016. Disponível em: <https://www.lincolninst.edu/publications/other/atlas-urban-expansion-2016-edition>.
- CABLA – Los Angeles’s Community Analysis Bureau. *The state of the city report: A cluster analysis of Los Angeles: a report*. Los Angeles, Community Analysis Bureau, 1974.
- CARTA Brasileira para Cidades Inteligentes. Brasília, 2021. Disponível em: <https://www.gov.br/cidades/pt-br/assuntos/desenvolvimento-urbano/carta-brasileira-para-cidades-inteligentes/CartaBrasileiraparaCidadesInteligentes2.pdf>.

- DIJKSTRA, L. et al. "A harmonised definition of cities and rural areas: the new degree of urbanization". *Regional Working Paper 2014*, WP 01/2014. European Commission Directorate General for regional and Urban Policy, 2014. Disponível em: https://ec.europa.eu/regional_policy/sources/work/2014_01_new_urban.pdf.
- IBGE – Instituto Brasileiro de Geografia e Estatística. *Censo demográfico 2022*. Rio de Janeiro, IBGE, 2023. Disponível em: <https://biblioteca.ibge.gov.br/visualizacao/livros/liv102011.pdf>.
- INTELI.GENTE. *Plataforma brasileira para diagnóstico de maturidade para cidades inteligentes e sustentáveis*. Brasília, MCTI, 2023. Disponível em: <https://inteligente.mcti.gov.br>.
- ISO – International Organization for Standardization. *Sustainable cities and communities – Indicators for urban services and quality of life*. Genebra, 2018. Disponível em: <https://www.iso.org/standard/68498.html>.
- ISO – International Organization for Standardization. *Sustainable cities and communities – Indicators for smart cities*. Genebra, 2019a. Disponível em: <https://www.iso.org/standard/69050.html>.
- ISO – International Organization for Standardization. *Sustainable cities and communities – Indicators for resilient cities*. Genebra, 2019b. Disponível em: <https://www.iso.org/standard/70428.html>.
- ITU – International Telecommunication Union. "Assessment framework for digital transformation of sectors in smart cities". SSCMM-ITU. ITU-T Y.4906. Genebra, 2019. Disponível em: <https://www.itu.int/rec/T-REC-Y.4906-201907-P>.
- ITU – International Telecommunication Union. "Use cases on implemented or evaluated smart sustainable city solutions based on ITU-T. Y.4900 recommendations series. ITU-T Y supplement 78, SERIES Y: global information infrastructure, internet protocol aspects, next-generation networks, internet of things and smart cities". ITU, 2023. Disponível em: https://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-Y.Sup78-202309-I!!PDF-E&type=items.
- MCTI – Ministério da Ciência, Tecnologia e Inovação; Iara – Inteligência Artificial Recriando Ambientes. *Manual de referência para coleta e metrificação de dados para os indicadores da plataforma inteli.gente: Transformação digital para as cidades brasileiras*. São Carlos, ICMC/USP, 2023.
- METROPOLITAN. Department of Ancient Near Eastern Art. "Uruk: the first city", in *Heilbrunn timeline of art history*. New York, The Metropolitan Museum of Art, 2003. Disponível em: http://www.metmuseum.org/toah/hd/uruk/hd_uruk.htm.
- NATH, M. et al. "A global-scale review of smart city practice and research focusing on residential neighbourhoods". *Habitat International*, v. 142, 2023. Disponível em: <https://doi.org/10.1016/j.habitatint.2023.102963>.
- TOWNSEND, A. M. *Smart cities: Big data, civic hackers, and the quest for a new utopia*. New York, W. W. Norton, 2013.
- UN-HABITAT. *What is a city*. UN-Habitat, 2020. Disponível em: https://unhabitat.org/sites/default/files/2020/06/city_definition_what_is_a_city.pdf.
- UNITED NATIONS. Department of Economic and Social Affairs (Population Division). *World urbanization prospects: The 2018 Revision (ST/ESA/SER. A/420)*. New York, United Nations, 2019. Disponível em: <https://population.un.org/wup/Publications/>. Acesso em: 9/jan./2024.

A inteligência artificial na pesquisa agrícola

*Antônio Mauro Saraiva
Fernando Santos Osório
André Freitas Colaço
Debora Pignatari Drucker
Eduardo Mario Mendiando
Fernando Elias Corrêa
Filipi Miranda Soares
José Paulo Molin
Marcos Roberto Benso
Patrícia Angélica Alves Marques
Roberto Fray da Silva
Sílvia Helena Galvão de Miranda
Willian França Costa
Alexandre Cláudio Botazzo Delbem*

resumo

A agricultura é um dos setores mais importantes da economia, fundamental para prover alimentos, fibras e energia para a humanidade. No Brasil é importante fonte de empregos e para uma balança comercial positiva. É também um dos setores que mais impactam o ambiente, pela mudança do uso do solo e pelas práticas de produção que podem resultar em poluição de solo, água e ar. Para fazer frente à complexidade dos seus processos de produção, a inteligência artificial tem desempenhado um papel cada vez mais importante na pesquisa agrícola, inovação e transformação social. Este texto aborda esse papel em algumas áreas importantes: gestão da água, produção agrícola e cadeias de valor, e a questão dos dados para a pesquisa.

Palavras-chave: inteligência artificial; agricultura; agricultura digital; tomada de decisão.

abstract

Agriculture is one of the most important sectors of the economy, essential for providing food, fiber and energy for humanity. In Brazil, it is an important source of jobs and a positive trade balance. It is also one of the sectors that most impacts the environment, due to changes in land use and production practices that can result in soil, water and air pollution. To face the complexity of its production processes, artificial intelligence has played an increasingly important role in agricultural research, innovation and social transformation. This text addresses this role in some important areas: water management, agricultural production and value chains, and the issue of data for research.

Keywords: *artificial intelligence; agriculture; digital farming; decision making.*

A

agricultura e a cadeia de produtos e serviços associados são de extrema importância no contexto econômico, social e ambiental no Brasil e no mundo. Diversos fatores impactam a produção agrícola, como o uso de insumos, emprego de novas tecnologias, questões e impactos ambientais, incluindo as mudanças climáticas. Eles afetam a produção de alimentos, sua disponibilidade e custo, e isso é central para a segurança alimentar. É necessário aumentar a produção agrícola, para atender às demandas atuais e de uma população crescente, mas diminuindo seus impactos ambientais, pois eles podem ter consequências na própria agricultura, por exemplo, pela perda de solo, diminuição dos recur-

sos hídricos, mudanças do clima e eventos extremos associados. De fato, o setor agrícola e de alimentos foi considerado como central para o futuro das questões ambientais e de saúde humana e para o futuro da saúde planetária, segundo a *São Paulo Declaration on Planetary Health*, elaborada pela comunidade global de saúde planetária (Myers; Pivor; Saraiva, 2021).

É, portanto, necessário e urgente que se desenvolvam novas abordagens para maximizar a produtividade e minimizar os impactos, para se ter uma produção agrícola sustentável, sem conflito com o meio ambiente.

A agricultura passou por grande evolução desde seus primórdios, e foi marcada pela incorporação de tecnologias e



Os créditos dos autores encontram-se no final do artigo.

inovações. Tem-se, inclusive, proposto denominações como Agricultura 1.0 até Agricultura 5.0 a períodos marcados por tecnologias específicas. Iniciando-se com os métodos mais rudimentares, como o cultivo manual e o uso de ferramentas simples, a agricultura ganhou grande impulso com a introdução de máquinas, como tratores e colhedoras, que aumentaram a capacidade de trabalho, a eficiência e a produtividade. A partir dos anos 1990 houve uma maior adoção de tecnologias da informação e comunicação (TICs), como GPSs e sensores, que viabilizaram a agricultura de precisão, permitindo um monitoramento mais preciso das lavouras e gestão mais eficiente dos recursos. A partir dos anos 2010, uma integração ainda maior entre TIC e agricultura ganha corpo e tecnologias como internet das coisas (IoT), big data e inteligência artificial levaram a uma agricultura mais conectada e automatizada, com drones para monitoramento, sistemas de irrigação inteligentes e análise de dados para tomada de decisões. Dissemina-se, então, a expressão “agricultura digital” e a evolução ocorre, também, no processamento de dados na lavoura, automação dos modelos de produção e de modelos preditivos, e outras inovações tecnológicas (Ahmad; Nabi, 2021).

É nesse ponto que a inteligência artificial tem grande impacto e enorme potencial, por ser capaz de auxiliar em diferentes etapas e processos da produção agrícola, melhorando a produtividade, eficiência e custos, minimizando impactos e aumentando a sustentabilidade. As diversas técnicas da IA permitem o desenvolvimento de sistemas inteligentes capazes de aprender e tomar decisões com base em dados, o que é fundamental na

agricultura. Possibilitam a identificação de padrões complexos em grandes conjuntos de dados, permitindo a previsão de eventos e a tomada de decisão mais eficiente.

Antes de chegar à produção agrícola, a IA tem sido crescentemente aplicada à pesquisa agrícola. A dimensão do setor agrícola e das pesquisas é tal que neste texto abordaremos alguns exemplos dessas aplicações, o suficiente para deixar claros a importância e o potencial a explorar.

IA NA PESQUISA AGRÍCOLA: ALGUNS EXEMPLOS

A questão dos dados agrícolas na pesquisa, desenvolvimento e inovação

O setor agrícola é multidisciplinar, envolve diversos setores da sociedade e é distribuído geograficamente. Isso acentua os desafios enfrentados na gestão de dados na pesquisa agrícola. A falta de padronização e a ineficiência na manipulação de dados (Saraiva et al., 2020) afetam a pesquisa agrícola, fundamental para a tomada de decisões informada e a inovação no agronegócio. De acordo com relatório da CrowdFlower (2016), cientistas de dados gastam cerca de 80% do tempo apenas coletando e organizando dados.

Os princípios FAIR (*findable, accessible, interoperable e reusable*) visam garantir que os dados sejam facilmente encontráveis, acessíveis, interoperáveis e reutilizáveis (Wilkinson et al., 2016). Ao facilitar o acesso, a padronização e o compartilhamento de dados também na

agricultura, os princípios FAIR facilitam o uso de IA e podem acelerar inovações, melhorar as práticas agrícolas e aumentar a sustentabilidade do setor.

No Brasil, a rede GO FAIR Agro Brasil¹ estimula a aplicação dos princípios FAIR em dados agrícolas, promovendo estratégias e parcerias, incentivando o compartilhamento e a reutilização de dados no âmbito dos sistemas de produção e na pesquisa agrícola (Drucker et al., 2021a).

Padrões de metadados, ontologias e *knowledge graphs* (KG) estão emergindo como recursos valiosos na gestão de dados agrícolas e são bons exemplos da aplicação dos princípios FAIR. Uma abordagem estruturada para o gerenciamento de dados no setor agrícola é importante, visando não só à melhoria na precisão e eficiência, mas também à facilitação do acesso e reutilização dos dados cruciais para a pesquisa.

Padrões de metadados são especialmente importantes nesse contexto, pois, ao fornecer descrições detalhadas e padronizadas dos dados, facilitam a localização e compreensão deles, permitindo assim que usuários e sistemas de IA identifiquem de forma eficaz o conteúdo, a origem e o contexto dos dados. Reer et al. (2023) destacam esses benefícios para *AI driven research* e Bakis et al. (2023) abordam a inclusão de metadados na anotação de imagens, com informações sobre os objetos na imagem, suas propriedades visuais, condição da amostra e propriedades técnicas do arquivo digital. Isso permite que cien-

tistas de *machine learning* (ML) filtrem, processem e reconheçam objetos em suas pesquisas. Isso é muito relevante na agricultura, em que o uso de imagens é intenso.

No contexto brasileiro, pesquisadores do Centro de Inteligência Artificial (C4AI-USP) e da Embrapa têm trabalhado no desenvolvimento de um esquema de metadados para a agricultura, o Agriculture and Livestock Metadata Elements Set (Almes Core) (Soares et al., 2022). O primeiro conjunto de metadados foi definido para descrever dados de comercialização agrícola. Na sequência, outros domínios da agricultura devem ser cobertos pelo esquema de metadados. Ele pode ser acessado no GitHub² e é aberto aos interessados na curadoria e evolução do esquema.

É importante destacar que, além dos princípios FAIR, a Food and Agriculture Organization (FAO) também destaca que a IA pode melhorar a governança de dados na segurança alimentar e segurança nutricional, desde que sejam seguidos e promovidos os princípios conhecidos como CARE (“benefício coletivo”, “autoridade para controlar a gestão de dados”, “responsabilidade”, “ética”) (FAO, 2023; FAO; ITPGRFA, 2023).

A prática dos dados abertos, porém, esbarra em questões culturais, políticas e mesmo em estratégias (Fairbairn; Kish, 2023). Dados do setor privado em geral enfrentam a desconfiança e o temor de vazamento e de abrir caminho para investigações fiscais ao quebrar segredos de opções estratégicas. Um contraexemplo

1 Disponível em: <https://go-fair-agro.github.io/>. Acesso em: 29/jan./2024.

2 Disponível em: <https://github.com/AlmesCore/AlmesCore>. Acesso em: 25/jan./2024.

positivo a destacar é o caso do Centro de Estudos Avançados em Economia Aplicada (Cepea) da USP³, em que a confiança das empresas na universidade superou essas barreiras. Elas compartilham dados confidenciais que são utilizados de modo agregado para a composição de importantes indicadores de mercado.

No setor público, apesar da existência da Lei de Acesso à Informação (LAI) (Brasil, 2011), que visa promover a transparência e facilitar o acesso a dados governamentais, ainda se observa uma forte resistência ao compartilhamento de dados. Essa resistência pode ser atribuída, em parte, ao receio de expor falhas metodológicas ou vulnerabilidades que possam comprometer de alguma forma as instituições ou órgãos envolvidos. Contudo, é essencial reconhecer que a exposição e o debate aberto sobre tais falhas poderiam ser encarados como oportunidades para aprimorar processos, aumentar a relevância dos dados e das instituições responsáveis, e promover uma cultura de melhoria contínua e transparência.

Gestão da água e mudanças climáticas

A agricultura é a principal consumidora de água do mundo, responsável por 87% do consumo global de água, e 60% de toda a retirada global de água doce vai para irrigação (Wu et al., 2022). A disponibilidade hídrica é essencial para

plantas e animais e influi diretamente na produtividade agrícola. Atender à demanda crescente de alimentos, limitando o aumento do consumo de água, considerando as mudanças climáticas, requer maior eficiência no uso da água e menor intensidade de consumo e desperdício. Pode-se destacar dois aspectos relevantes sobre a gestão da água para a produção agrícola sustentável: planejamento e gestão integrada.

O primeiro é o planejamento para lidar com o potencial impacto de eventos climáticos extremos – secas ou excesso de chuvas, responsáveis por 18% a 43% da variabilidade da produtividade agrícola (Vogel et al., 2019), sendo as secas um problema mais pronunciado e principal responsável por perdas em produtividade em 75% das áreas agrícolas do mundo (Kim; Iizumi; Nishimori, 2019). O segundo refere-se à gestão integrada dos recursos hídricos, que inclui o uso eficiente de irrigação e a melhoria da resiliência da produção não irrigada (Liu et al., 2022).

Silva et al. (2021c) utilizaram aprendizagem não supervisionada para identificar o potencial impacto de secas e diferentes zonas de produtividade, utilizando informações climáticas, de produtividade e espaciais. Um estudo de caso com cana-de-açúcar conclui que o uso do framework proposto proporcionou melhores resultados que o método tradicional de avaliação pelo Standard Precipitation Index (SPI), que indica seca meteorológica.

Visto que dados distribuídos e heterogêneos são um problema frequente na agricultura, Silva et al. (2021b) desenvolveram um framework teórico para organizar os estudos e a pesquisa, que permite a

3 Disponível em: <https://www.cepea.esalq.usp.br/br/sobre-o-cepea.aspx>. Acessado em: 5/mar./2024.

avaliação de cenários de mudanças climáticas e o auxílio à tomada de decisões por diferentes agentes. Adota conceitos de gerenciamento de riscos e o uso tanto de modelos climáticos quanto de IA, permitindo a extração de conhecimento de diferentes tipos de dados em contextos distintos. Aspectos de modelagem multiobjetivos também são considerados pelos autores.

O planejamento para lidar com extremos climáticos engloba três pontos: (i) o desenho de planos de ação para os diferentes cenários; (ii) a montagem de um sistema de monitoramento, como estações meteorológicas próximas à área agrícola; e (iii) a execução do plano de ação em questão, considerando a automação de tarefas.

A pesquisa em modelos e técnicas de IA auxilia a prover potenciais soluções para os três pontos citados. No caso dos planos de ação, a pesquisa utilizando IA foca atividades como: auxiliar na identificação de padrões nas séries climáticas; auxiliar no entendimento do impacto de diferentes ações passadas no manejo de recursos naturais; e auxiliar na tomada de decisões sobre irrigação em dado período. O desperdício também é uma questão bastante relevante, combinando a irrigação com os dados de umidade do solo e de expectativa/previsão de chuvas.

Em relação à execução do plano de ação em questão, a IA entra como recurso adicional em pesquisas sobre automatização inteligente de sistemas de irrigação, monitoramento do impacto da seca nos talhões, utilizando drones, e monitoramento no nível de solo, utilizando redes de sensores sem fio e IoT.

Balti et al. (2020) realizam uma ampla revisão bibliográfica que destaca a IA

como um importante conjunto de métodos que pode auxiliar na extração de conhecimento e auxílio à tomada de decisões, relacionados ao monitoramento de secas utilizando dados em grandes volumes.

Rolnick et al. (2022) apresentam uma revisão aprofundada do uso de IA na pesquisa em diferentes setores para endereçar problemas relacionados às mudanças climáticas. Para os sistemas agropecuários e florestais, apontam os seguintes usos: sensoriamento remoto de emissões de poluentes, agricultura de precisão, monitoramento de uso do solo e gestão de florestas.

Com relação à pesquisa em IA voltada ao correto uso da irrigação, uma série de fatores é considerada na literatura, como: (i) a quantidade de água; (ii) a qualidade da água disponível; (iii) os potenciais impactos de mudanças climáticas; (iv) a automação do sistema de irrigação; e (v) os impactos econômicos e sociais.

No que tange à quantidade e à qualidade de água disponível e os potenciais impactos de mudanças climáticas neste aspecto, a pesquisa em IA engloba: previsão de impactos de mudanças climáticas; previsão de vazão de rios; realização de simulações considerando diferentes cenários; realização de estudos para avaliar a sustentabilidade dos sistemas de irrigação; potenciais impactos do uso da água disponível nas plantas e nos sistemas de irrigação, considerando sua qualidade.

Tung e Yaseen (2020) analisam diferentes métodos para avaliação de qualidade da água em rios utilizando modelos de IA, e destacam o uso de redes neurais, lógica *fuzzy* e modelos híbridos, pois têm melhor desempenho que os modelos clássicos. Obaideen et al. (2022) e Garcia et al.

(2020) revisam a literatura do uso de IA e sistemas inteligentes em irrigação, destacando as principais tecnologias envolvidas, suas limitações e o importante papel da IA no desenvolvimento de novas práticas.

Produção agrícola: sustentabilidade, produtividade, eficiência

Na produção agrícola, os insumos e as práticas culturais, devidamente integrados, resultam em conversão de biomassa direcionada principalmente para a produção de alimentos, fibra e energia. Para atender ao setor, três dos principais objetivos buscados pela pesquisa agrícola são produtividade, sustentabilidade e eficiência, essenciais para atender à demanda crescente por diferentes tipos de alimentos oriunda de aumento populacional, melhoria da qualidade de vida e nível de renda, e demanda por produtos mais sustentáveis.

A IA na pesquisa agrícola pode auxiliar em diversos aspectos para aumentar a produtividade, com melhorias em: identificação de melhorias em processos; automação; aplicação e distribuição de insumos; processo de tomada de decisões, dentre outros. Outra contribuição é o potencial de aumento da qualidade dos produtos por melhorias no tratamento de doenças e pragas, na manipulação dos produtos, estratégias de manejo para maior qualidade nutricional e redução de perdas em campo. Vários desses pontos contribuem para a sustentabilidade.

Estratégias de otimização das lavouras baseadas na agricultura de precisão (AP) têm sido propostas e testadas, com

resultados impactantes, por exemplo, na produção e utilização de insumos (Shafi et al., 2019; Colaço et al., 2020). Com o avanço das TICs e o uso da ciência de dados no setor agrícola, pode-se estruturar o arcabouço de conhecimento agrônomo e de informações digitais para se obter recomendações de campo mais assertivas. Os sistemas inteligentes para o auxílio à tomada de decisões baseados em inteligência artificial podem contribuir de forma significativa.

Modelos de IA baseados em aprendizagem supervisionada, não supervisionada e por reforço são aplicados na pesquisa para aumento de produtividade, sustentabilidade e eficiência. A aprendizagem supervisionada tem sido usada na predição da produtividade. Há exemplos para diferentes culturas, como em Maldaner et al. (2021) e Canata et al. (2021), para cana-de-açúcar; Martello et al. (2022) e Bazame et al. (2022), para café; e Wei et al. (2020), para cenoura. Também tem sido utilizada para pesquisa na previsão de ocorrência de doenças e identificação de plantas daninhas (Liakos et al., 2018).

A aprendizagem não supervisionada pode ser utilizada para explorar os dados gerados no campo, auxiliando na análise de agrupamento de dados espaciais e temporais, na identificação de diferentes níveis de qualidade de produtos, dentre outros. Uma área de pesquisa importante em AP é a identificação de regiões, chamadas de zonas de manejo, que sejam mais homogêneas no tempo e no espaço para variáveis do solo e para produtividade. Busca-se encontrar padrões espaciais e temporais a partir de diversas camadas de dados, de inúmeras variá-

veis e com um grande número de pontos (dados). Zang et al. (2009) foram alguns dos pioneiros. Anselmi et al. (2021) utilizam zonas de manejo identificadas para adotar diferentes níveis de população de plantas por área, de acordo com a aptidão de cada região. Em Zane et al. (2013) e Costa (2016), métodos de agrupamentos são combinados com análise de autocorrelação espacial para obter segmentos mais significativos aos objetivos desejados.

A aprendizagem por reforço tem aplicação em pesquisas relacionadas à execução de tarefas de forma automática ou autônoma, como nos robôs agrícolas e na automação de sistemas. Uma importante aplicação na pesquisa é na experimentação *on-farm*, ou “experimentação de precisão”, baseada em testes em campo em larga escala, implementados por meio de ferramentas de automação e monitores com sensores. Gera uma grande quantidade de dados sobre a resposta da cultura ao tratamento de interesse e sobre as características locais que explicam tal resposta. Difere da experimentação agrícola tradicional, em pequenas parcelas e com alto controle estatístico. Busca o auxílio à tomada de decisão por meio de observações empíricas a campo, dada uma alteração de manejo (Lacoste et al., 2022; Bramley et al., 2022). A automação desse processo e o uso de técnicas avançadas de tratamento e análise de dados no âmbito da IA oferecem grandes oportunidades para sistemas inteligentes de auxílio à tomada de decisão, como em Colaço et al. (2021). Colaço et al. (2024) demonstraram essa abordagem, com um modelo de aprendizado de máquina que foi capaz de prever a dose ótima de

fertilizante nitrogenado a ser aplicada, relacionando as características de campo observadas a condições para as quais a dose ótima era conhecida, por meio de um banco de dados histórico oriundo da experimentação *on-farm*. Essa abordagem superou o desempenho (em termos de lucratividade e acurácia da recomendação) de qualquer outro método baseado em conhecimento agrônomico empregado tradicionalmente na produção agrícola.

IA nas pesquisas em cadeias de valor agrícolas

São inúmeros os exemplos do potencial de uso e de impacto dos diferentes instrumentos da inteligência artificial na pesquisa nas cadeias de valor e comercialização agropecuária, e em particular em segurança alimentar.

Há mais de 30 anos, o Cepea-USP⁴ realiza levantamentos de preços diários em todas as regiões do Brasil, dependendo da cultura. Para os dados agropecuários as variáveis de tempo e localização são importantes para buscar relações, fazer cruzamentos e visualizar em mapas ou outros formatos. Grandes volumes de dados históricos têm sido armazenados de forma permanente e aplicados em modelagem de séries temporais, analisando questões como a formação e transmissão de preços e a relação de troca de produtos agropecuários, tanto em nível de pesquisa como para apoio direto ao mercado.

4 Disponível em: <https://www.cepea.esalq.usp.br/br>. Acesso em: 25/jan./2024.

O uso cada vez mais amplo de IA, como em algoritmos de aprendizagem de máquina, mineração de dados e textos, e ainda a organização de grandes bases de dados, beneficia fortemente a pesquisa. Aplicar modelos de redes socioeconômicas permite avaliar padrões de transporte de cargas, de movimentação de animais e vegetais, de exportações, que podem subsidiar políticas e programas de fomento à produção, de controle de doenças e pragas, de suporte de preços a regiões mais vulneráveis e de sistemas de transporte. Como exemplo, Menezes, Luna e Miranda (2020) aplicam o modelo de redes socioeconômicas utilizando registros do Guia de Trânsito Animal (GTA) de bovinos, emitidos pelos órgãos de defesa sanitária para autorização de circulação de animais nos estados e entre estados, para avaliar o padrão geográfico da movimentação, para fins de discussão de política de defesa sanitária, com foco na doença da febre aftosa.

A IA nos estudos de cadeias de valor auxilia na categorização de informações de forma ágil e com suporte estatístico, e permite identificar padrões em conjuntos de dados quantitativos e também em bases de informações textuais. É o caso da mineração de textos para analisar o impacto de notícias ambientais sobre fluxos comerciais do Brasil para o exterior (Oliveira; Miranda, 2023; Silva, 2022).

Também nas análises de risco no trânsito internacional de produtos e, conseqüentemente, na definição de padrões amostrais em comércio de produtos de origem vegetal e animal, a IA é importante aliada, permitindo poupar tempo e recursos e a construção de sistemas como o do canal-verde na inspeção de embalagens e suportes de madeira no

aeroporto de Viracopos, em Campinas. É o caso dos estudos para a definição do sistema atualmente em funcionamento pela concessionária ABV no aeroporto de Viracopos, com fiscalização da madeira pelo Vigiagro/Mapa (Adami; Miranda; Marcondes, 2021). Técnicas de ML foram usadas para identificar o padrão de comportamento de empresas e, a partir desses modelos, ter mais elementos para definir sistemas de fiscalização, com aplicações possíveis do ponto de vista sanitário e fiscal (Aranha, 2021).

Benso et al. (2023a) englobam diferentes cadeias de valor no estudo de modelos e metodologias para o desenho e avaliação de seguros agrícolas considerando múltiplas ameaças climáticas. Concluem o trabalho com um exemplo para a produção de soja no Brasil, utilizando aprendizagem não supervisionada para melhor explorar, analisar e entender os dados relacionados a perdas, produtividade e índices climáticos relevantes.

Silva et al. (2021a) usam aprendizagem supervisionada na pesquisa relacionada a preços de produtos agrícolas. Exploraram o uso de modelos clássicos de previsões de séries temporais (Arima e Sarima), de modelos de IA (SVR, AdaBoost e LSTM) e de conjuntos destes modelos (*ensembles*) para previsão de preços diários de soja e açúcar no Brasil. Concluíram que os modelos com melhores resultados para os dois produtos foram os de IA e que os clássicos foram os piores.

O Instituto Nacional de Ciência e Tecnologia de Combate à Fome⁵, criado em 2023, traz explicitamente a IA como um

5 Disponível em: <https://www.fsp.usp.br/inct-combate-a-fome/> Acesso em: 25/jan./2024.

eixo transversal a todas as pesquisas. Ele surgiu da interação entre pesquisadores de diversas áreas, incluindo o eixo AgriBio⁶ do C4AI-USP. A área tem potencial e demanda enormes atualmente, dado que a fome e a insegurança alimentar se agravaram.

Na segurança alimentar, a IA pode auxiliar a construir indicadores de oferta e demanda de alimentos, a partir de dados detalhados, visando-se obter um mapa de vulnerabilidade que permite avaliações com cruzamento de diferentes indicadores. O cruzamento de dados de saúde pública com dados geográficos, sociais, econômicos, de gastos públicos, de localização de equipamentos da política pública, entre outros, é possível e pode ser muito útil para o planejamento, implantação e monitoramento de resultados das políticas e programas públicos para combate à insegurança alimentar, como no trabalho de Gomes e Miranda (2023).

Domene et al. (2023) discutem temas importantes sobre segurança alimentar, englobando diversos aspectos da pesquisa e do uso de IA para auxiliar o desenvolvimento de políticas públicas e a tomada de decisão de diferentes agentes. Apontam que a ciência de dados e a IA são essenciais para auxiliar a enfrentar desafios relacionados à fome e à insegurança alimentar.

DISCUSSÃO E CONCLUSÃO

Não há como não reconhecer ou subestimar a importância da IA para a pesquisa

e, na sequência, para a produção e comercialização agrícola. As características dos processos agrícolas, sua abrangência, a complexidade natural dos aspectos biológicos, climáticos e de solo, ambientais e de sustentabilidade, a diversidade de perfis de produtores, as instabilidades de mercado – do local ao global –, entre outros fatores, requerem ferramentas adequadas para permitir seu entendimento e a tomada de decisões bem-informada, muito mais do que ocorre hoje.

Os exemplos acima são uma pequena amostra do que existe, e ainda menor se considerarmos a demanda e o potencial. A agricultura traz desafios para a pesquisa em IA, por exemplo, pela necessidade de melhores modelos (previsibilidade/confiabilidade) e de explicabilidade para que os produtores possam entender e aceitar as recomendações e os resultados. Como os dados são um ponto de partida, a questão dos princípios FAIR e CARE, juntos, torna-se crítica. Destaque para a dificuldade de acesso aos dados, seu compartilhamento e qualidade. Quando sob a custódia de entes públicos, o acesso é muitas vezes limitado por burocracia e políticas internas de não disponibilização, apesar da vigência da Lei de Acesso à Informação. Nesse sentido, programas como o Open Government Partnership, ou Parceria para Governo Aberto, iniciativa internacional da qual o Brasil foi um dos fundadores, são muito importantes, pois estimulam os entes governamentais a adotarem práticas de abertura e compartilhamento de dados. Quando de fontes privadas, o receio de desvelar dados que possam ter implicações fiscais, por exemplo, limita o seu compartilhamento,

6 Disponível em: https://c4ai.inova.usp.br/research_2/#AgriBio_B_eng. Acesso em: 25/jan./2024.

que poderia ser útil para entender o que se passa em cada local. Para dar mais segurança aos demais atores, é crucial o desenvolvimento de políticas sobre IA e uso de dados.

É necessária atuação multidisciplinar na pesquisa em IA na agricultura, pelas características dos seus problemas, atuação esta que já consta formalmente de políticas institucionais nos meios acadêmicos, seja nas universidades, nos órgãos de avaliação, seja ainda nas agências de fomento. A realidade, porém, é que a prática está longe de estimular a multidisciplinaridade, tanto nas avaliações dos pesquisadores, suas carreiras, sua produção, como na avaliação de projetos, que muitas vezes acabam nas mãos de avaliadores que, com uma ótica monodisciplinar, não reconhecem o valor e as características de propostas e times multidisciplinares.

Em linha com isso está a necessidade de atuar em conjunto com os demais *stakeholders*, desde a identificação das questões até o desenho e execução de projetos de pesquisa, em processo de cocriação.

A inteligência artificial desempenhará um papel cada vez mais importante na agri-

cultura, integrada com outras tecnologias como IoT, e poderá proporcionar *insights* mais profundos, possibilitando decisões mais precisas e informadas no campo e ao longo das cadeias produtivas, tornando a agricultura mais inteligente e sustentável.

O uso da IA demandará mais infraestrutura de dados e processamento para sua implementação nas pesquisas na agricultura. A fragmentação dos recursos por projetos não é uma boa solução, pois não permite otimizar seu uso e dificulta obter a escala necessária. Por outro lado, as universidades brasileiras têm capacidade muito limitada de atender institucionalmente a essa demanda, caso prevaleçam as políticas e os investimentos vigentes.

Nesse contexto, oportunidades a serem exploradas são: parcerias público-privadas, que possuem um benefício adicional de troca de experiências e conhecimento e facilitam a transferência de tecnologia; o desenho e implantação de soluções de infraestrutura compartilhadas entre diferentes centros de pesquisa; e programas de fomento de longo prazo, que englobem tanto o investimento inicial em infraestrutura quanto a sua manutenção e modernização.

REFERÊNCIAS

- ADAMI, A. C. O.; MIRANDA, S. H. G.; MARCONDES, A. G. "Fiscalização de embalagens de madeira em Viracopos". *Revista de Política Agrícola*, v. 30, 2021, pp. 45-53.
- AGGARWAL, M.; KHULLAR, V.; GOYAL, N. "Agriculture in society 5.0", in V. Khullar et al. (eds.). *Artificial intelligence and society 5.0: issues, opportunities and changes*. Londres, Chapman and Hall/CRC, 2024.
- AHMAD, L.; NABI, F. *Agriculture 5.0: artificial intelligence, IoT and machine learning*. Boca Raton, CRC Press, 2021.
- ANSELMINI, A. A. et al. "Definition of optimal maize seeding rates based on the potential yield of management zones". *Agriculture*, v. 11, 2021, pp. 911-27.
- ARANHA, F. S. *Programa de fiscalização reduzida em embalagens de madeira para acondicionamento de importados: impacto do programa no comportamento das empresas e comparação entre modelos preditivos para o manejo do risco*. Dissertação de mestrado. Piracicaba, Escola Superior de Agricultura Luiz de Queiroz da USP, 2021.
- BALTI, H. et al. "A review of drought monitoring with big data: issues, methods, challenges and research directions". *Ecological Informatics*, v. 60, 2020, pp. 101-36.
- BAZAME, H. C. et al. "Mapping coffee yield with computer vision". *Precision Agriculture*, v. 23, 2022, pp. 1-16.
- BENSO, M. R. et al. "Design and evaluation of weather index insurance for multi-hazard resilience and food insecurity". *Natural Hazards and Earth System Sciences*, v. 23, n. 4, 2023a, pp. 1.335-54.
- BRAMLEY, R. G. V. et al. "Did someone say 'farmer-centric'? Digital tools for spatially distributed on-farm experimentation". *Agronomy for Sustainable Development*, v. 42, n. 105, 2022.
- BRASIL. LEI nº 12.527, de 18 de novembro de 2011. *Diário Oficial da União*, seção 1, edição extra, p. 1. Disponível em: <https://www.lexml.gov.br/urn/urn:lex:br:federal:lei:2011-11-18;12527>.
- CANATA, T. F. et al. "Sugarcane yield mapping using high-resolution imagery data and machine learning technique". *Remote Sensing*, v. 13, 2021, pp. 232-46.
- COLAÇO, A. F. et al. "How will the next-generation of sensor-based decision systems look in the context of intelligent agriculture? A case-study". *Field Crops Research*, v. 270, fev./2021, pp. 108-205.
- COLAÇO, A. F. et al. "Digital strategies for nitrogen management in grain production systems: lessons from multi-method assessment using on-farm experimentation". *Precision Agriculture*, 2024. Disponível em: <https://link.springer.com/article/10.1007/s11119-023-10102-z>.
- COLAÇO, A. F. et al. "Economic viability, energy and nutrient balances of site-specific fertilisation for citrus". *Biosystems Engineering*, v. 200, 2020, pp. 138-56. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1537511020302488?via%3Dihub>.
- COSTA, W. F. *Segmentação multirresolução variográfica ótima*. Tese de doutorado. São Paulo, Poli-USP, 2016.
- CROWDFLOWER. *Data science report*. CrowdFlower, 2016. Disponível em: <https://www2.cs.uh.edu/~ceick/UDM/CFDS16.pdf>. Acesso em: 5/set./2023.

- DOMENE, S. et al. "Segurança alimentar: reflexões sobre um problema complexo". *Estudos Avançados*, v. 37, 2023, pp. 181-206.
- DRUCKER, D. P. et al. "Implantação da rede temática GO FAIR Agro Brasil: primeiros passos". *Anais do XIII Congresso Brasileiro de Agroinformática*, pp. 164-71. Porto Alegre, 2021. Disponível em: <https://sol.sbc.org.br/index.php/sbiagro/article/view/18387>.
- FAIRBARN, M.; KISH, Z. "Setting data free: the politics of open data for food and agriculture". *New Media & Society*, v. 25, n. 8, 2023, pp. 1.935-59. Disponível em: <https://journals.sagepub.com/doi/10.1177/14614448231174520>.
- FAO. "CFS policy recommendations on strengthening FSN data collection and analysis tools for food security and nutrition". 2023. Disponível em: https://www.fao.org/fileadmin/templates/cfs/Docs2223/Data/OEWG-23-03/CFS_Policy_Recommendations_on_FSN_Data_ZeroDraft_final_Mar16.pdf.
- FAO; ITPGRFA. "Developments regarding digital sequence information/genetic sequence data". Rome, may/2023. Disponível em: <https://www.fao.org/3/cc5400en/cc5400en.pdf>.
- GARCÍA, L. et al. "IoT-based smart irrigation systems: an overview on the recent trends on sensors and IoT systems for irrigation in precision agriculture". *Sensors*, v. 20, n. 4, 2020, pp. 10-42.
- GOMES, J. G.; MIRANDA, S. H. G. "Mapeando a vulnerabilidade à insegurança alimentar na metrópole paulista". XI Encontro Nacional da Anpaas. Curitiba, 12-15/set./2023.
- KIM, W.; IIZUMI, T.; NISHIMORI, M. "Global patterns of crop production losses associated with droughts from 1983 to 2009". *Journal of Applied Meteorology and Climatology*, v. 58, n. 6, 2019, pp. 1.233-44.
- LACOSTE, M. et al. "On-farm experimentation for global agriculture transformation". *Nature Food*, v. 3, 2022, pp. 11-18. Disponível em: <https://www.nature.com/articles/s43016-021-00424-4>.
- LIAKOS, K. G. et al. "Machine learning in agriculture: a review". *Sensors*, v. 18, n. 8, 2018, 1-29. Disponível em: <https://www.mdpi.com/1424-8220/18/8/2674>.
- LIU, X. et al. "Global agricultural water scarcity assessment incorporating blue and green water availability under future climate change". *Earth's Future*, v. 10, n. 4, 2022.
- MALDANER, L. F. et al. "Predicting the sugarcane yield in real-time by harvester engine parameters and machine learning approaches". *Computers and Electronics in Agriculture*, v. 181, 2021.
- MARTELLO, M. et al. "Coffee-yield estimation using high-resolution time-series satellite images and machine learning". *Agriengineering*, v. 4, 2022, pp. 888-902.
- MENEZES, T. C. de; LUNA, I.; MIRANDA, S. H. G. de. "Network analysis of cattle movement in Mato Grosso do Sul (Brazil) and implications for foot-and-mouth disease". *Frontiers in Veterinary Science*, v. 7, 2020, p. 219.
- MYERS, S. S.; PIVOR, J. I.; SARAIVA, A. M. "The São Paulo Declaration on Planetary Health". *The Lancet*, v. 398, n. 10.308, 2021, pp. 12-99.
- OBAIDEEN, K. et al. "An overview of smart irrigation systems using IoT". *Energy Nexus*, 2022, pp. 100-24.
- OLIVEIRA, M. M.; MIRANDA, S. H. G. "A importância da reputação no comércio internacional de commodities agrícolas brasileiras". *Anais do 61º Congresso da Sociedade Brasileira de Economia, Administração e Sociologia Rural (Sober)*. Piracicaba, Even3, 2023.

- REER, A. et al. "FAIR human neuroscientific data sharing to advance AI driven research and applications: legal frameworks and missing metadata standards". *Frontiers in Genetics*, v. 14, 1086802. Disponível em: <https://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2023.1086802/ful>.
- ROLNICK, D. et al. "Tackling climate change with machine learning". *ACM Computing Surveys (CSUR)*, v. 55, n. 2, 2022, pp. 1-96.
- SARAIVA, A. M. et al. "Dados na agricultura digital: ciclo, padronização, qualidade, compartilhamento e segurança", in D. M. Queiroz et al. (eds.). *Agricultura digital*. Viçosa, Universidade Federal de Viçosa, 2020, pp. 308-25.
- SHAFI, U. et al. "Precision agriculture techniques and practices: from considerations to applications". *Sensors*, v. 19, n. 17, 2019, pp. 37-96.
- SILVA, J. M. *A interface entre meio ambiente e comércio internacional: uma análise para o comércio agrícola brasileiro com a União Europeia*. Dissertação de mestrado. Piracicaba, Escola Superior de Agricultura Luiz de Queiroz da USP, 2022.
- SILVA, R. F.; BARREIRA, B. L.; CUGNASCA, C. E. "Prediction of corn and sugar prices using machine learning, econometrics, and ensemble models". *Engineering Proceedings*, v. 9, n. 1, 2021a, p. 31.
- SILVA, R. F. et al. "A theoretical framework for multi-hazard risk mapping on agricultural areas considering artificial intelligence, IoT, and climate change scenarios". *Engineering Proceedings*, v. 9, n. 1, 2021b, p. 39.
- SILVA, R. F. et al. "A data-driven framework for identifying productivity zones and the impact of agricultural droughts in sugarcane using SPI and unsupervised learning". 2021 IEEE International Workshop on Metrology for Agriculture and Forestry (MetroAgriFor), 2021c, pp. 226-31.
- SIMEONAKI, E.; KRISTANOS, M.; ARVANITIS, K. "Recent advances in digital twins for agriculture 5.0: applications and open issues in livestock production systems". *Applied Sciences*, v. 14, n. 2, 2024.
- SOARES, F. M. et al. "Building a community-based fair metadata schema for Brazilian agriculture and livestock trading data". *CEUR Workshop Proceedings*, v. 3.235, 2022.
- TUNG, T. M.; YASEEN, Z. M. "A survey on river water quality modelling using artificial intelligence models: 2000-2020". *Journal of Hydrology*, v. 585, pp. 124-670.
- VOGEL, E. et al. "The effects of climate extremes on global agricultural yields". *Environmental Research Letters*, v. 14, n. 5, 2019.
- WEI, M. C. F. et al. "Carrot yield mapping: a precision agriculture approach based on machine learning". *AI*, v. 1, 2020, pp. 229-41.
- WILKINSON, M. D. et al. "The FAIR guiding principles for scientific data management and stewardship". *Scientific Data*, 3:160018, 2016. Disponível em: <https://www.nature.com/articles/sdata201618>.
- WU, B. et al. "Quantifying global agricultural water appropriation with data derived from Earth observations". *Journal of Cleaner Production*, v. 358, 2022. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0959652622015013>.

ANTÔNIO MAURO SARAIVA é professor do Departamento de Engenharia de Computação da Escola Politécnica e do Centro de Inteligência Artificial (C4AI), ambos da USP.

FERNANDO SANTOS OSÓRIO é professor do Instituto de Ciências Matemáticas e de Computação (ICMC) da USP e do C4AI-USP

ANDRÉ FREITAS COLAÇO é professor do Departamento de Engenharia de Biossistemas da Escola Superior de Agricultura Luiz de Queiroz (Esalq) da USP.

DEBORA PIGNATARI DRUCKER é analista da Embrapa Agricultura Digital e colaboradora do C4AI-USP.

EDUARDO MARIO MENDIONDO é professor da Escola de Engenharia de São Carlos da USP e coordenador do Centro de Estudos e Pesquisas em Desastres (Ceped) e da Cátedra Unesco de Águas Urbanas.

FERNANDO ELIAS CORRÊA é pesquisador do C4AI-USP e do Centro de Estudos Avançados em Economia Aplicada (Cepea) da USP.

FILIPI MIRANDA SOARES é doutorando na USP e na Universidade de Twente (Holanda) e pesquisador do C4AI-USP.

JOSÉ PAULO MOLIN é professor do Departamento de Engenharia de Biossistemas da Esalq-USP e pesquisador do C4AI-USP.

MARCOS ROBERTO BENSO é bolsista CNPq no INCT ONSEAdapta e pesquisador do C4AI-USP.

PATRICIA ANGÉLICA ALVES MARQUES é professora do Departamento de Engenharia de Biossistemas da Esalq-USP e do C4AI-USP.

ROBERTO FRAY DA SILVA é professor do Departamento de Engenharia de Biossistemas da Esalq-USP e pesquisador do C4AI-USP.

SÍLVIA HELENA GALVÃO DE MIRANDA é professora do Departamento de Economia, Administração e Sociologia da Esalq-USP e pesquisadora do Cepea-USP e do C4AI-USP.

WILLIAN FRANÇA COSTA é pós-doutorando na Escola Politécnica da USP.

ALEXANDRE CLÁUDIO BOTAZZO DELBEM é professor do ICMC-USP e do C4AI-USP.



**Ética em IA e IA ética:
prolegômenos e estudo
de casos significativos**

Luís C. Lamb

resumo

A inteligência artificial é uma tecnologia de propósito geral que tem causado mudanças significativas nas atividades econômicas, com implicações em todas as atividades da vida humana. Como é característico das tecnologias disruptivas, diversas inovações decorrentes delas, bem como as suas consequências sobre a humanidade, causam impactos que vão além das atividades de base tecnológica. Em particular, a compreensão dos aspectos éticos inerentes ao desenvolvimento de tecnologias baseadas em IA apresenta desafios distintos de experiências anteriores. Neste trabalho, apresentamos uma introdução objetiva aos principais conceitos de ética que podem ser alinhados ao desenvolvimento da IA, além de delinear desafios inerentes ao desenvolvimento de sistemas que apresentem comportamento ético delineado de acordo com os valores e princípios éticos consensualizados pela sociedade.

Palavras-chave: ética em IA; inteligência artificial; aprendizado de máquina; IA ética.

abstract

Artificial intelligence is a general purpose technology that has caused significant changes in economic activities, with implications in all human life aspects. As is characteristic of disruptive technologies, various innovations resulting from them, as well as their consequences on humanity, cause impacts that go beyond technology-based activities. In particular, understanding the ethical aspects inherent to the development of basic technologies in AI presents different challenges from previous experiences. In this work, we present an objective introduction to the main ethical concepts that can be aligned with the development of AI. In addition, we highlight challenges inherent to the development of systems designed and aligned with ethical behavior outlined in accordance with the ethical values and principles agreed upon by society.

Keywords: *ethics in AI; artificial intelligence; machine learning; ethical AI.*

A

inteligência artificial é hoje considerada uma tecnologia de propósito geral que tem causado implicações e disrupções mensuráveis sobre todos os aspectos da vida humana. Este amplo uso, a atenção obtida entre setores significativos da população e o impacto da IA nas últimas duas décadas sugeriram aos pesquisadores a preocupação quanto a aspectos éticos e riscos da área, que podem ir além dos resultados científicos e tecnológicos (Donnelly, 2017). Hoje, a IA transforma diretamente a economia, a educação, as relações de trabalho em diversas escalas, as relações de poder e a (geo)política internacionais (Burton, 2017; Bostrom; Yudkowsky, 2014; Brynjolfsson; Rock; Syverson, 2018). Natu-

ralmente, devido às implicações multissetoriais da IA – como é característico de toda tecnologia de propósito geral –, pesquisadores e pensadores vinculados a esses outros domínios, além da ciência da computação, passam a investigar de forma mais sistemática as consequências das transformações causadas sobre a humanidade (Brynjolfsson; Rock; Syverson, 2018; Lipsey; Carlaw; Bekar, 2006).

Já no início do século XXI, a partir do impacto público da IA, primeiramente na academia e posteriormente causado pelas empresas de tecnologia, o físico Stephen Hawking (Universidade de Cambridge), assim como os empresários de tecnologia Bill Gates (cofundador da Microsoft) e

LUÍS C. LAMB é professor da Universidade Federal do Rio Grande do Sul (UFRGS).

Elon Musk (CEO da Tesla), expressaram e anteciparam publicamente suas preocupações sobre o impacto (inclusive ético e moral) da IA sobre a vida humana, o futuro do trabalho e da sociedade (Sainato, 2015). Nesse mesmo período, iniciaram-se, por parte de organizações científicas como a Association for the Advancement of Artificial Intelligence (AAAI), a Association for Computing Machinery (ACM), a Royal Society e o Institute of Electrical and Electronic Engineers (IEEE), entre diversas outras, a organização de workshops, conferências e grupos de trabalho sobre os impactos éticos da IA e sobre como construir IA sob o prisma da ética (Donnelly, 2017; Anderson; Anderson; Armen, 2005; Walsh, 2015; Furman et al., 2018.).

Historicamente, no entanto, desde a organização dos primeiros workshops, conferências e publicações acadêmicas sobre IA, um número reduzido de pesquisadores se dedicava à análise e investigação da temática da ética em tecnologia ou ética das máquinas¹. *Ética em IA* e *IA ética*, por sua vez, não eram temas centrais amplamente investigados ou de preocupação principal entre os pesquisadores da área de IA. Dados acerca desta não preocupação entre os pesquisadores líderes em IA no mundo acadêmico foram

sintetizados por Prates, Avelar e Lamb (2018)². Igualmente, esse estudo foi utilizado como base para o influente *AI Index Report*, edições de 2019 e 2021, publicado anualmente pela Universidade de Stanford sob coordenação do HAI – Stanford Institute for Human-Centered Artificial Intelligence (Mishra; Clark; Perrault, 2020; Perrault et al., 2019; Zhang et al., 2021).

Neste breve artigo, faremos um panorama inicial sobre como as duas temáticas da ética em IA e da IA ética se conceitualizam e se relacionam atualmente

1 A primeira conferência sobre a temática da IA ocorreu na Universidade de Dartmouth, em 1956 (denominada “Dartmouth Summer Research Project on Artificial Intelligence”), organizada por pioneiros da área como John McCarthy, Marvin Minsky, Nathaniel Rochester e Claude Shannon. Allen Newell e Herbert Simon (Nobel de Economia em 1978) também participaram do encontro. Ver, para aspectos históricos da IA: Audibert et al. (2023).

2 Nesse estudo, para avaliar o quanto a ética na IA é analisada nos artigos científicos das principais conferências e revistas de IA, os autores identificaram os termos relacionados à ética nos títulos de artigos em conferências e periódicos de referência em IA, aprendizado de máquina e robótica publicados entre 1969 e 2017. O percentual de palavras-chave tem uma interpretação simples e direta. Para cada categoria de terminologia da palavra-chave buscada (IA clássica, tendência atual e ética) identifica-se o número de artigos para o qual o título, ou resumo do artigo, contém pelo menos uma palavra dentre as pesquisadas. No caso das conferências anuais da Association for the Advancement of AI (AAAI) ou da Neural Information Processing Systems (NeurIPS), identificam-se as palavras-chave no resumo do artigo. Pode haver classificação de um artigo em mais de uma categoria, por exemplo, um artigo que contenha uma palavra-chave clássica e uma palavra-chave classificada como ética. Os termos analisados basearam-se nas questões expostas e identificadas em trabalhos de referência na área de IA e também nos temas para discussão na First AAAI/ACM Conference on AI, Ethics and Society. As palavras-chave utilizadas foram as seguintes: *accountability, accountable, employment, ethic, ethical, ethics, fool, fooled, fooling, humane, humanity, law, machine bias, moral, morality, privacy, racism, racist, responsibility, rights, secure, security, sentience, sentient, society, sustainability, unemployment* e *workforce* (como as publicações em IA são realizadas na língua inglesa, preservei aqui o original das palavras-chave). Os conjuntos de palavras-chave clássicas e de tendência foram compilados a partir das áreas do livro mais citado sobre IA, de Stuart Russell e Peter Norvig (2020), e da curadoria de termos das palavras-chave que apareceram com mais frequência em títulos de artigos ao longo do tempo nas conferências e revistas analisadas. As análises foram realizadas sobre um conjunto de mais de 100 mil artigos.

dentro do contexto de pesquisa em IA, notadamente na academia e empresas de tecnologia, bem como ilustraremos algumas iniciativas globais que visam garantir o uso ético da inteligência artificial. O trabalho não é uma revisão sistemática ou revisão bibliográfica, apenas tem como objetivo oferecer a pesquisadores e ao público leigo uma conceitualização útil e referências que podem expandir a análise desses temas centrais em pesquisa e desenvolvimento na área de inteligência artificial. Este trabalho é direcionado ao público amplo, não apenas a especialistas. Assim, apresentaremos conceitos básicos para entendimento do texto ao longo do mesmo. Sugere-se, como literatura inicial sobre o estado da arte em IA, os textos publicados neste dossiê da **Revista USP**, o livro de Marcus e Davis (2019) e o capítulo recentemente publicado em língua portuguesa (Bazzan et al., 2023), que cobre os principais avanços da IA em termos de pesquisa na última década. Não cobriremos os aspectos da regulamentação da IA neste trabalho, embora a temática evidentemente se relacione aos aspectos, principalmente deontológicos, da ética em IA.

ÉTICA NA IA E IA ÉTICA: BREVE CONCEITUALIZAÇÃO

A inteligência artificial é o domínio da ciência no qual pesquisadores analisam, desenvolvem e experimentam – sob a tutela do método científico – como construir sistemas e tecnologias que apresentem de forma mensurável habilidades ou funções cognitivas, como o raciocínio e o aprendizado de máquinas. Há inúmeras

definições de IA, mas para o escopo deste trabalho esta nos é suficiente. Outros conceitos subjacentes se referem à *IA forte* (que potencialmente replicaria a cognição humana em múltiplas tarefas), *IA fraca* (que se refere a sistemas que apresentam um único foco de aplicação – como identificar imagens de uma única categoria, como animais domésticos ou um tradutor automático simples) e *IA geral* (no inglês: *artificial general intelligence* – AGI), que se refere à construção de tecnologias ou máquinas que tenham a capacidade de aplicar soluções inteligentes para problemas de qualquer natureza³.

A *ética na IA*, um ramo da ética aplicada (assim como a bioética), recebeu atenção recente de pesquisadores devido aos potenciais *riscos* acerca de uma tecnologia de propósito geral (IA) potencialmente capaz de alterar significativamente a vida humana (Burton, 2017; Bostrom; Yudkowsky, 2014; Brynjolfsson; Rock; Syverson, 2018; Lipsey; Carlaw; Bekar, 2006; Sainato, 2015; Marcus; Davis, 2019; Bazzan, 2023; Anderson; Anderson, 2011). Aliado a estas preocupação e motivação, o desenvolvimento de tecnologias e produtos computacionais em geral e sistemas de IA em particular passou a receber especial atenção. Pesquisadores e lideranças

3 Mesmo entre especialistas, as distinções entre esses conceitos não são universalmente aceitas. Outros conceitos como *HAI* (*human-level AI*) e *superinteligência artificial* foram propostos. *Human-level AI* (inteligência no nível humano) se refere a sistemas que teriam as habilidades cognitivas humanas, capacidade de autoaprendizado e autonomia, como o ser humano. A *superinteligência artificial* teria habilidades cognitivas superiores à humana. No entanto, enfatizamos que esses conceitos ainda possuem intersecções e incertezas em suas definições.

globais alertam para a necessidade de construirmos sistemas que tenham como requisitos o uso ético da tecnologia. Neste caso, nos referimos à IA ética – aquela tecnologia ou produto que faz uso intensivo da IA que necessariamente passe a apresentar valores morais ou que se comporte de acordo com a moralidade (e valores) consensualizados pela sociedade.

Aqui, antes de prosseguirmos na análise, temos conceitos que precisam ser definidos para melhor leitura. Uma conceitualização geral, adotada neste trabalho, define ética como o estudo e análise dos fenômenos morais. A análise de valores morais que o ser humano ou a sociedade seguem sob o paradigma rigoroso da investigação filosófica recai sobre a ética normativa, que no caso da IA e das ciências remete à ética aplicada. O estudo da *metaética* analisa as (meta)definições dos conceitos fundamentais da área, como valores e julgamentos morais. Investiga-se, em ética normativa, aquilo que em lógica (filosófica) se associa a *modalidades* como: o *permitido* (em termos de comportamento humano, social ou das “máquinas”, incluindo a IA); o *obrigatório* (que tipicamente passa a ser regulado por leis ou regulamentos nas organizações, sociedades e nações); as *proibições* (tipicamente consensualizadas pelas organizações e sociedades e também reguladas por leis, normatizadas ou regulamentadas); e também se formalizam as *omissões* (aquilo que não é obrigatório ou consensualizado dentro um sistema normativo).

Esta análise, do ponto de vista mais rigoroso, remete à deontica no prisma da lógica filosófica, que pode ser relacionada a formulações através de um sistema denominado pelos lógicos e filósofos

como lógicas modais. Em lógicas modais estudamos afirmações como “é necessário que” ou “é possível que” como qualificadores de uma afirmação ou proposição (Åqvist, 1994; Gabbay, 2013). A formalização padrão da lógica deontica é descrita formalmente em lógicas modais no estilo de Kripke, isto é, no qual as proposições são interpretadas através dos modelos de mundos possíveis; ou interpretações aceitas, no caso da lógica deontica⁴. Na lógica deontica as modalidades representam, por exemplo, obrigações, permissões e proibições. No entanto, a formalização rigorosa ou lógica destes conceitos vai além do escopo deste trabalho⁵. Ainda assim, é útil mencionar a literatura básica sobre lógica deontica ao leitor, para partirmos da mesma conceitualização ou ontologia.

OS DESAFIOS DA ÉTICA NA IA E DA CONSTRUÇÃO DE TECNOLOGIAS ÉTICAS DE IA

Como mencionamos acima, a preocupação com aspectos éticos da IA não era vista ou mensurada como prioritária até o início do século XXI. Alguns trabalhos, como o de Prates, Avelar e

4 Brevemente: sob uma interpretação ao estilo de Kripke, dizemos que uma afirmação (em lógica, chamamos afirmação de *proposição*) é obrigatória se em todos os mundos possíveis (aceitos) ou em todas as interpretações esta proposição tem de ser verdadeira. Isto é, uma obrigação tem de ser seguida ou obedecida sob todas as interpretações. Entretanto, a noção de *permissão* não implica a noção de obrigação (em todos os mundos possíveis ou interpretações aceitas). Essas noções são elegantemente formalizadas em Åqvist (1994) e Gabbay (2013).

5 Ver Åqvist (1994) e Gabbay (2013) para uma análise formal da lógica deontica.

Lamb (2018), conforme relatado no *AI Index Report* de 2019 e 2021 (Russell; Norving, 2003; Mishra; Clark; Perrault, 2020; Perrault et al., 2019), identificaram que um percentual abaixo de um por cento dos trabalhos publicados nas principais conferências de IA e aprendizado de máquina, entre 1969 e 2017, se referia a aspectos relacionados ao impacto ético dos resultados científicos e tecnológicos apresentados nesses trabalhos, pelo menos de forma mais explícita nos textos dos artigos. Ainda na linha de identificação e classificação de trabalhos científicos que tinham como foco as implicações éticas da IA, Avelar, Audibert e Lamb (2022) apresentaram uma metodologia que utiliza IA justamente para classificar se um determinado artigo aborda aspectos éticos. Nesse trabalho, os autores utilizaram uma base de dados rotulada por especialistas como conjunto de treinamento e avaliaram uma ampla base de artigos para realizar a classificação, mostrando, pelo menos em princípio, como a metodologia utilizada na IA pode ser utilizada para identificar estudos que lidam com aspectos éticos.

Questões fundamentais quanto à ética em (e da) inteligência artificial vão certamente além das preocupações dos pesquisadores e organizações quanto à avaliação inicial do impacto das tecnologias de IA expostas nas suas publicações (Avelar; Audibert; Lamb, 2022). Na década de 2010, trabalhos pioneiros na área de ética identificaram diversos riscos do uso indiscriminado ou pelo menos não integralmente avaliado, normatizado e mensurado da IA. Esses riscos e consequências incluíram desafios que a construção de sistemas éticos de IA tem a responder.

Para que o leitor tenha uma noção concreta e realista dos desafios enfrentados na construção de IA ética, bem como da necessidade de integração de princípios éticos na IA, descrevemos quatro estudos de casos reais que ocorreram recentemente. Um desses trabalhos trata de um projeto de pesquisa visando responder a desafios inerentes ao uso da IA na seleção de pessoas para posições de trabalho. De forma inovadora, esse projeto remete à construção de uma investigação sobre questões relevantes para o mundo do trabalho por uma equipe multidisciplinar de potencial de alto impacto, liderado por pesquisadores da Universidade de Harvard. A seguir, descrevemos os casos em que se tornou explícita a necessidade de tratarmos a ética na IA e o uso da IA ética de forma sistematizada.

- Vieses de gênero, raça e nacionalidade em tecnologias de tradução automática e processamento de linguagem natural. Prates, Avelar e Lamb (2020) demonstraram que a ferramenta de tradução automática mais utilizada no mundo, Google Translate, pode exibir vieses de gênero e uma forte tendência para *defaults*/padrões masculinos, embora os autores tenham demonstrado que esses vieses resultam dos dados do mundo real utilizados no treinamento do sistema de tradução automática e há uma provável relação com a forma como a sociedade fala e escreve sobre gênero no mundo do trabalho. Através da análise sistemática da tradução de frases como “Ele/Ela é um engenheiro” (e outras similares) de línguas neutras de gênero, como o húngaro e chinês,

para o inglês, os autores coletaram estatísticas sobre a assimetria entre os pronomes de gênero feminino e masculino nos resultados da tradução. Mostraram, também, que os *defaults* masculinos, além de proeminentes, são exacerbados em áreas com estereótipos, como Stem (*science, technology, engineering and mathematics*). No artigo, os resultados também ressaltaram que o fenômeno do viés de gênero em tradução automática iria além das questões das profissões: a proporção de pronomes femininos variava significativamente nos experimentos realizados à época, de acordo com os adjetivos utilizados para descrever uma pessoa. Por exemplo, adjetivos como *shy* (envergonhada/o) e *desirable* (desejável) foram traduzidos em larga proporção e associados com pronomes femininos, enquanto adjetivos como *guilty* (culpado) e *cruel* (cruel) foram quase exclusivamente traduzidos para o masculino. Posteriormente a esse trabalho, foi percebido um notável desenvolvimento de uma linha de pesquisa em ética em IA e ética em processamento de linguagem natural (PLN), na qual pesquisadores passaram a analisar vieses em sistemas de PLN que utilizam múltiplas línguas e mecanismos de tradução, como, por exemplo, Fan (2021), Sheng (2019) e Devinney e Björklund (2022).

- Vieses na seleção de pessoas por departamentos de recursos humanos que utilizam as *hiring platforms* ou plataformas tecnológicas de contratação de pessoal (Kenthapadi; Venkataraman, 2017; Harvard, 2023).

O trabalho de pesquisa liderado pela professora da Universidade de Harvard

Cynthia Dwork (Harvard, 2023) parte das premissas de que sistemas hoje utilizados em departamentos de recursos humanos, embora sejam justos quando utilizados em isolamento, não são necessariamente justos em seu conjunto. O trabalho consiste fundamentalmente de uma equipe multidisciplinar que adotará uma abordagem holística para identificar problemas reais nesses tipos de sistemas de IA, hoje amplamente adotados nas organizações. Especificamente, sistemas cujas atividades de trabalho são descritas insuficientemente podem levar à exclusão de candidatos qualificados, bem como a própria linguagem utilizada pelo candidato pode levar à sua exclusão por conter vieses culturais.

Nesse sentido, o projeto irá “explorar IA e outras técnicas que as plataformas de contratação (*hiring platforms*) podem utilizar para permitir uma comparação imparcial entre os candidatos”. Outro desafio atual é que as plataformas que utilizam IA para contratação usam algoritmos de ranqueamento e pontuação de candidatos a empregos que acentuam as injustiças. Os pesquisadores desse projeto liderado por Dwork irão desenvolver e pesquisar algoritmos que anulem esses efeitos e melhorem o desempenho dos sistemas de IA utilizados em contratação de pessoas. Outros pontos a serem investigados nessa linha de pesquisa multidisciplinar envolvem a utilização de algoritmos preditivos para calibrar os escores preditivos do sucesso das pessoas em uma determinada posição de trabalho quando contratada. Para isso, os pesquisadores fazem uso da teoria

da pseudo-aleatoriedade para “questionar abordagens que assegurem que os modelos preditivos são tão acurados quanto possível”. Também será pesquisada nesse projeto a busca de evidências sobre como modelar um ambiente de trabalho igualitário – como as plataformas de contratação que usam IA podem tornar os mecanismos de recomendações pessoais e digitais e as redes de relacionamento mais justas. Finalmente, outro ponto importante desse projeto multidisciplinar será a identificação de marcos regulatórios legais que podem vir a avançar a equidade das plataformas tecnológicas de contratação de pessoas. Em suma, esse projeto ilustra que as abordagens multidisciplinares podem ser uma resposta apropriada aos desafios que enfrentamos hoje quanto a desenvolver sistemas éticos de IA.

- Vieses raciais em algoritmos de IA que classificaram pessoas como não sendo seres humanos (Garcia, 2016; BBC News, 2015).

Conforme relatado em 2015, Jacky Alcine, enquanto utilizava o sistema conhecido como Google Photos, percebeu que o aplicativo de reconhecimento de rostos (*faces*) rotulava ele e seu amigo como gorilas. Posteriormente, Alcine postou uma foto do sistema no Twitter que imediatamente se tornou viral nas redes sociais. Essa notícia causou grande repercussão na época, com veículos jornalísticos de todo o mundo reportando o assunto, incluindo a BBC News (2015). Além dos danos e prejuízos pessoais que podem ser múltiplos e cuja análise aprofundada vai além do escopo deste trabalho, este

evento ocorrido ainda em 2015 ilustra a repercussão que os impactos éticos da IA podem ter sobre os usuários e, também, sobre organizações. No caso, uma empresa especializada em tecnologia de informação e pioneira em diversas aplicações da IA teve de se responsabilizar pelas consequências.

Também levantamos a necessidade da reflexão, tanto por parte de profissionais e pesquisadores de IA quanto por equipes multidisciplinares sobre a relevância dos impactos éticos dessas tecnologias: nesse sentido, é relevante o exemplo do projeto relatado no item 2. Também nos parece claro que os indivíduos que fazem uso inadvertido ou desinformado dessas tecnologias podem vir a ser impactados, sendo esta ainda uma temática com múltiplas questões em aberto. Afinal, sobre quem recairá a responsabilidade sobre os impactos pessoais da utilização da IA? Sobre as empresas que produzem as tecnologias? Sobre os pesquisadores e engenheiros que desenvolveram os sistemas? Ou sobre os múltiplos agentes envolvidos na produção, certificação, padronização e posterior manutenção das tecnologias? Essas perguntas remetem ao conceito da IA responsável, que abordaremos sucintamente na última seção. Também é relevante mencionar que as iniciativas regulatórias dos países e organismos multilaterais visam compreender as melhores formas de responder a essas questões.

- Implicações do uso ético de tecnologias de IA por planos de saúde.

No trabalho altamente citado de Obermeyer et al. (2019), os autores mos-

traram que planos de saúde que utilizam algoritmos preditivos de IA e aprendizado de máquina exibiram vieses raciais significativos. Esse é outro desafio e problema da IA ética, pois o uso de tecnologias na saúde pode afetar milhares de potenciais pacientes. Os pesquisadores demonstraram no seu artigo, publicado na revista *Science*, que o viés do sistema utilizado “surge porque o algoritmo prevê os custos dos cuidados de saúde e não a doença, mas o acesso desigual aos cuidados significa que gastamos menos dinheiro a cuidar de pacientes negros do que de pacientes brancos”. Argumentam também que, “apesar dos custos dos cuidados de saúde parecerem ser um substituto eficaz para a saúde por algumas medidas de precisão preditiva, surgem grandes preconceitos raciais”. Esse foi um trabalho pioneiro na área, demonstrando que aspectos éticos na IA, combinados com sistemas que devem fazer uso de valores éticos na sua construção, são fundamentais para a resolução de vieses e preconceitos que afetam uma população muito significativa.

Notadamente, como as abordagens atuais da IA são fundamentadas fortemente em bases de dados específicas para treinamento dos algoritmos de aprendizado de máquina, a utilização de bases de dados que não representam parcelas significativas e diversas da população mundial, combinada com práticas de desenvolvimento de sistemas que não consideram valores éticos, pode levar a consequências imprevistas e, principalmente, a sistemas que apresentam altos riscos para a população.

CONCLUSÃO: NA DIREÇÃO DA IA ÉTICA, CONFIÁVEL E RESPONSÁVEL

Este artigo apresentou uma breve introdução a temáticas inter-relacionadas e relevantes para o desenvolvimento da inteligência artificial ética. Abordamos aspectos relacionados à ética na IA, que se propõem a analisar como integrar valores morais consensualizados pelas sociedades nos sistemas tecnológicos de IA, bem como aspectos referentes à IA ética, que se trata do desenvolvimento de sistemas que demonstrem intenção de atuarem moral e eticamente, de acordo com os valores da sociedade. A abordagem utilizada foi a apresentação dos conceitos fundamentais que permitam ao leitor não especializado compreender os quatro estudos brevemente descritos. Esses estudos de caso ilustram os desafios atuais que pesquisadores e profissionais da área de IA enfrentam no desenvolvimento de tecnologias que incorporem valores éticos e sejam consideradas confiáveis pela sociedade.

Nesse sentido, cabe aqui apontar algumas tendências atuais de áreas de pesquisa que visam responder aos desafios pertinentes à integração entre ética e inteligência artificial. Um desses desafios diz respeito à construção de várias formas de IA que sejam consideradas *trustworthy AI*, *responsible AI* e *ethical AI*. Desse modo, um número significativo de pesquisadores se debruça sobre como construir inteligência artificial *confiável*, *responsável* e *ética*, respectivamente. Esses conceitos se inter-relacionam e apresentam desafios em comum para pesquisadores que atuam na área de IA.

Visando à construção de sistemas confiáveis, responsáveis e éticos, as pesquisas atuais apontam ser necessário responder aos desafios da *explicabilidade, interpretabilidade, semântica e responsabilidade*. Os avanços recentes na IA generativa⁶, da qual tecnologias como ChatGPT (da empresa OpenAI) e LLaMA (da empresa Meta) são exemplos⁷, acentuam ainda mais os desafios da construção de sistemas de IA confiáveis e éticos. Tais desafios são significativos pois esses sistemas não apresentam um modelo semântico formal que explique o seu funcionamento. Isto é, esses sistemas são considerados caixas-pretas com ausência de explicabilidade ou interpretabilidade. Isto significa – por mais que o leitor se surpreenda – que os pesquisadores ainda não explicaram rigorosamente se as respostas produzidas por esses sistemas estão corretas ou se são produzidas com alguma garantia de correção. Por exemplo, quando um usuário interage com grandes modelos de linguagem (Wolfram, 2023), como ChatGPT,

6 A IA generativa se refere a tecnologias de IA que produzem (geram) textos, imagens, músicas e outras expressões que representam conhecimento a partir de modelos de redes neurais artificiais (notadamente os modelos baseados em *transformers*; veja uma explicação simples no artigo de Stephen Wolfram [2023], bem como no artigo técnico de Rachel Gordon [2023]), que manipulam e aprendem a gerar aproximações sobre um grande conjunto de dados. Os sistemas se tornaram recentemente populares pela explosão de interesse dos usuários, notadamente pela tecnologia (*chatbot*) conhecida como ChatGPT. Este *chatbot* foi desenvolvido pela empresa de tecnologia OpenAI e lançado em novembro de 2022. O ChatGPT foi desenvolvido a partir das versões 3.5 e 4.0 da tecnologia chamada GPT (*generative pre-trained transformer*). Informações adicionais podem ser obtidas em: <https://chat.openai.com/>.

7 Grandes modelos de linguagem – do inglês *large language models* (LLMs).

as respostas obtidas muitas vezes são denominadas de “alucinações”, ou seja, são criadas pelo sistema mas não têm nenhuma relação semântica com a pergunta do usuário. Outro fenômeno observado por pesquisadores foi a apresentação de vieses de gênero por parte dos grandes modelos de linguagem (Gordon, 2023).

Para responder a esses desafios, pesquisadores propuseram o uso de abordagens que combinem e integrem as abordagens de IA baseadas em redes neurais (que são eficientes para aprendizado sobre um grande volume de dados) com as abordagens da IA que utilizam a lógica simbólica (que permite a formulação rigorosa do processo de raciocínio). Esses sistemas, denominados *inteligência artificial neurosimbólica* (Garcez; Lamb, 2023), permitem, em princípio, interpretar e explicar de forma lógica e rigorosa o comportamento dos sistemas de IA construídos sob a abordagem das redes neurais artificiais. Isso decorre do fato de sistemas lógicos, por construção, terem semântica rigorosamente definida, além de estabelecerem os fundamentos também rigorosos da inferência e do raciocínio e por permitirem a representação formal de normas e valores, que podem ser representados nos sistemas de IA, idealmente mais éticos.

Em conclusão, é notório que os atuais sistemas de aprendizado de máquina (*machine learning*) se constituem no principal avanço da IA nas últimas duas décadas. Esses sistemas de aprendizado são a base fundamental das tecnologias da IA generativa, dos tradutores automáticos de linguagem, classificadores de imagens e fotografias, entre outras aplicações. Eles propiciaram a explosão da adoção

das tecnologias de IA por organizações públicas, empresas e pela sociedade, com repercussão global e sobre estratégias de desenvolvimento de países e implicações em relações bilaterais, tendo em vista a relevância da IA na economia.

No entanto, os sistemas atuais de IA construídos sobre algoritmos baseados em redes neurais artificiais são, como já dito, considerados caixas-pretas pelos pesquisadores. Isso significa que não há ainda na literatura científica explicações rigorosas sobre o que essas caixas-pretas aprendem ou computam, para sermos mais precisos (Garcez; Lamb, 2023). A utilização de sistemas de IA que não sejam explicáveis ou interpretáveis pode vir a aumentar os riscos decorrentes do uso indiscriminado, inadvertido ou desinformado por parte de indivíduos, organizações e, até mesmo,

por parte de governos. Para mitigar e responder a esses desafios, a construção de IA ética se apresenta como uma linha de pesquisa fundamental na academia. As pesquisas que visam identificar e analisar os princípios éticos para a construção de IA responsável, confiável e ética ainda permanecem como um grande desafio dos nossos tempos. Entretanto, tendo em vista o interesse atual por esse tema, confiamos nos resultados de pesquisas multidisciplinares na área. Nessa linha, também poderemos vislumbrar a posterior adoção desses princípios e valores éticos por profissionais e organizações de IA. Tais princípios éticos, imaginamos que talvez se tornem indispensáveis para o correto exercício profissional, bem como para a segura disseminação de tecnologias e produtos de IA em prol da humanidade.

REFERÊNCIAS

- ANDERSON, M.; ANDERSON, S. L. (eds.). *Machine ethics*. Cambridge, Cambridge University Press, 2011.
- ANDERSON, M.; ANDERSON, S. L.; ARMEN, C. *Machine ethics – Technical report*. FS-05-06. 2005 AAI Fall Symposium. AAI Press, Menlo Park, 2005.
- ÅQVIST, L. “Deontic logic”, in D. Gabbay; F. Guentner, (eds.). *Handbook of philosophical logic. Volume II – Extensions of classical logic*. Dordrecht, Kluwer, 1994.
- AUDIBERT, R. B. et al. “On the evolution of AI and machine learning: towards a meta-level measuring and understanding impact, influence and leadership at premier AI conferences”. *Journal of Applied Logics – The IFCOLOG Journal of Logics and their Applications*, v. 10, n. 5, 2023, pp. 693-817.

- AVELAR, P. H. C.; AUDIBERT, R. B.; LAMB, L. C. "Measuring ethics in AI with AI: a methodology and dataset construction". *BRACIS*, n. 1, 2022, pp. 370-84.
- BAZZAN, A. L. C. et al. "A nova eletricidade': aplicações, riscos e tendências da IA moderna", in *Escola de Computação PPGC/UFRGS 50 Anos*, 2023, pp. 167-209.
- BBC NEWS. "Google apologises for photos app's racist blunder". *BBC News*, 1/jul./2015. Disponível em: <https://www.bbc.com/news/technology-33347866>.
- BOSTROM, N.; YUDKOWSKY, E. "The ethics of artificial intelligence", in K. Frankish; W. M. Ramsey (eds.). *The Cambridge handbook of artificial intelligence*. Cambridge, Cambridge University Press, 2014, pp. 316-34.
- BRYNJOLFSSON, E.; ROCK, D.; SYVERSON, C. "The productivity j-curve: How intangibles complement general purpose technologies". *NBER Working Paper*, n. 25148, oct./2018. Revised jan./2020.
- BURTON, E. et al. "Ethical considerations in artificial intelligence courses". *AI Magazine*, v. 38, n. 2, 2017, pp. 22-34.
- DEVINNEY, H.; BJÖRKLUND, J.; BJÖRKLUND, H. "Theories of gender", in *NLP bias research*. Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, 2022.
- DONNELLY, P. et al. *Machine learning: the power and promise of computers that learn by example*. The Royal Society, 2017.
- FAN, A. et al. "Beyond English-centric multilingual machine translation". *Journal of Machine Learning Research*, v. 22, n. 107, 2021, pp. 1-48.
- FURMAN, J. et al. *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*. New Orleans, 2-3/feb./2018, ISBN 978-1-4503-6012-8.
- GABBAY, D. et al (eds.). *Handbook of deontic logic and normative systems*, v. 1. Londres, College Publications, 2013.
- GARCEZ, A. d'A.; LAMB, L. C. "Neurosymbolic AI: the 3rd wave". *Artificial Intelligence Review*, v. 56, n. 11, 2023.
- GARCIA, M. "Racist in the machine: the disturbing implications of algorithmic bias". *World Policy Journal*, v. 33, n. 4, 2016, pp. 111-7.
- GORDON, R. "Large language models are biased. Can logic help save them?". *MIT News*, 3/mar./2023. Disponível em: <https://news.mit.edu/2023/large-language-models-are-biased-can-logic-help-save-them-0303>.
- HARVARD John A. Paulson School of Engineering and Applied Science News. "How can bias be removed from artificial intelligence-powered hiring platforms? Harvard-led institute to pursue fairness in online systems". 12/jun./2023. Disponível em: <https://seas.harvard.edu/news/2023/06/how-can-bias-be-removed-artificial-intelligence-powered-hiring-platforms>.
- KENTHAPADI, K.; LE, B.; VENKATARAMAN, G. "Personalized job recommendation system at LinkedIn: practical challenges and lessons learned", in *Proceedings of the eleventh ACM conference on recommender systems*. 2017, pp. 346-7.
- LIPSEY, R.; CARLAW, K.; BEKAR, C. "Economic transformations: general purpose technologies and long-term economic growth". *The Economic History Review*, v. 59, n. 4, 2006, pp. 881-2.
- MARCUS, G.; DAVIS, E. *Rebooting AI: building artificial intelligence we can trust*. Vintage, 2019.
- MISHRA, S.; CLARK, J.; PERRAULT, C. R. "Measurement in AI policy: opportunities and challenges". *CoRR*, 2020.

- OBERMEYER, Z. et al. "Dissecting racial bias in an algorithm used to manage the health of populations". *Science*, v. 366, n. 6464, 2019, pp. 447-53.
- PERRAULT, R. et al. *The AI index 2019 annual report*. AI Index Steering Committee, Human-Centered AI Institute, Stanford University, dec./2019.
- PRATES, M. O. R.; AVELAR, P. H. C.; LAMB, L. C. "Assessing gender bias in machine translation: a case study with Google Translate". *Neural Comput. Appl.*, v. 32, n. 10, 2020, pp. 6.363-81.
- PRATES, M. O. R.; AVELAR, P. H. C.; LAMB, L. C. *On quantifying and understanding the role of ethics in AI research: a historical account of flagship conferences and journals*. Global Conference on AI (GCAI), 2018, pp. 188-201.
- RUSSELL, S.; NORVING, P. *Artificial intelligence: a modern approach*. Pearson, 2003.
- SAINATO, M. "Stephen Hawking, Elon Musk, and Bill Gates warn about artificial intelligence". *The Observer*. 19/ago./2015. Disponível em: <https://observer.com/2015/08/stephen-hawking-elon-musk-and-bill-gates-warn-about-artificial-intelligence/>.
- SHENG, E. et al. "The woman worked as a babysitter: on biases in language generation". *EMNLP/IJCNLP*, n. 1, 2019.
- VASWANI, A. et al "Attention is all you need". *NIPS*, 2017.
- WALSH, T. "Artificial intelligence and ethics". *2015 AAAI Workshop*. Austin, AAAI Press, 2015.
- WOLFRAM, S. "What is ChatGPT doing... and why does it work?". *Wolfram Media*, 2023.
- ZHANG, D. et al. "The AI index 2021 annual report". *CoRR*, 2021.



**Elas na inteligência
artificial – Questões de gênero**

Renata Wassermann

resumo

Há muitos anos se discute a baixa representatividade de alguns grupos nas áreas tecnológicas, principalmente mulheres e pessoas negras. Neste artigo, examinamos o caso da inteligência artificial, em que o panorama não é diferente. Discutimos os efeitos da falta de diversidade nos grupos que desenvolvem a IA e iniciativas ao redor do mundo para tentar mitigá-los. Mostramos um caso específico – sistemas de tradução automática, uma subárea da IA – em que os estereótipos de gênero aparecem.

Palavras-chave: IA e gênero; diversidade; viés de gênero; viés de raça; tradução e viés.

abstract

For many years, the low representation of some groups in technological areas, especially women and black people, has been discussed. In this article, we examine the case of artificial intelligence, where the panorama is not different. We discuss the effects of a lack of diversity in groups developing AI and initiatives worldwide to mitigate them. We show a specific case – machine translation systems, a sub-area of AI – in which gender stereotypes appear.

Keywords: AI and gender; diversity; gender bias; race bias; translation and bias.

Nos anos 80 e início dos anos 90, o termo “inteligência artificial” era quase sinônimo de sistemas especialistas, baseados em regras que codificavam de forma simbólica, através de alguma linguagem formal, o conhecimento de especialistas, como médicos ou químicos. No início do século 21, com o avanço das abordagens não simbólicas, baseadas em aprendizado de máquina, passamos a ver referências a “uma inteligência artificial”, ou à “IA da empresa X”, tratando cada sistema como um agente dotado de inteligência. Em 2022, com a popularização dos grandes modelos de linguagem como o GPT e o Bard, a IA passou a ser vista pelo público leigo como sendo apenas esta subárea de pesquisa. Os modelos de linguagem

são baseados em redes neurais profundas e aprendizado autossupervisionado, técnicas que já existem há décadas. No entanto, a escala e o escopo dos modelos aumentaram enormemente. Por exemplo, o GPT-3 tem 175 bilhões de parâmetros e pode ser adaptado por meio de instruções em linguagem natural para fazer uma ampla gama de tarefas, apesar de não ter sido treinado explicitamente para realizar muitas delas (Bommasani et al., 2022).

A primeira falta de diversidade que surge com a popularização dos grandes modelos é a de linhas de pesquisa. Rapidamente, a atenção e o financiamento se voltaram à construção de modelos de linguagem cada

RENATA WASSERMANN é professora associada do Departamento de Computação do Instituto de Matemática e Estatística (IME) da USP e pesquisadora do C4AI e do Lawgorithm, também da USP.

vez maiores, que para funcionar necessitam cada vez mais de recursos computacionais e fontes de dados para treinamento. Com isso, cada vez menos instituições possuem os recursos necessários, fazendo com que um número bastante pequeno de *big techs* domine o mercado. Países e continentes inteiros ficam excluídos do desenvolvimento desse tipo de inteligência artificial.

Outro grande problema é a amplificação de preconceitos e vieses presentes nas bases de dados utilizadas para treinamento desses modelos. Bommasani et al. (2022) descrevem diferentes tipos de vieses encontrados nos grandes modelos de linguagem. Apesar de serem problemas que aparecem em outros sistemas de inteligência artificial e ciências em geral, como os modelos são mal compreendidos e utilizados em diferentes aplicações, os danos são potencialmente maiores, como já ficou claro desde o lançamento do livro de Cathy O’Neil (2017).

Um caso bastante estudado atualmente é o de sistemas com viés de raça. Danos podem ir desde inconvenientes, como o dispenser de sabão automático que viralizou na internet por apenas reconhecer mãos de pessoas brancas, até carros que não reconhecem pedestres negros¹ e sistemas que caracterizam negros como criminosos ou mais propensos a reincidir em crimes. Buolamwini (2017) verificou sistematicamente ferramentas de reconhecimento facial e mostrou como as bases de dados usadas pelas ferramentas eram pouco diversas e funcionavam melhor para homens brancos. Muitas vezes, as pessoas nem sabem o que está sendo levado em

conta na sua avaliação, como no caso recente das universidades americanas². É importante lembrar que nos Estados Unidos pessoas de origem latina não são consideradas brancas e também sofrem discriminação.

De um modo geral, os vieses nos dados refletem preconceitos presentes na sociedade em que os dados foram criados. Como disse Benjamin (2019, p. 60)³, não é necessário ter má intenção, basta não se preocupar em como o passado molda o presente. Se tradicionalmente cozinhar era uma ocupação feminina, existem mais imagens de mulheres do que homens e, assim, sistemas passam a reconhecer pessoas na cozinha como mulheres⁴.

Parece bastante claro que algo precisa ser feito para mitigar os danos causados por vieses (Boinodiris, 2024), mas mesmo assim propostas de ações afirmativas ainda encontram bastante resistência, com alegações de que são discriminatórias e desnecessárias⁵ (West; Whittaker; Crawford, 2019). Podemos pensar em três aspectos de diversidade:

- diversidade dentro dos grupos que fazem a IA – se as pessoas que trabalham no desenvolvimento de sistemas de IA são brancas, do gênero masculino, podem passar despercebidas falhas como uso de estereótipos e linguagem inadequada;

1 Ver: <https://www.businessinsider.com/self-driving-cars-worse-at-detecting-dark-skin-study-says-2019-3>.

2 Ver: <https://themarkup.org/machine-learning/2021/03/02/major-universities-are-using-race-as-a-high-impact-predictor-of-student-success>.

3 “No malice needed, no N-word required, just lack of concern for how the past shapes the present.”

4 Ver: https://brasil.elpais.com/brasil/2017/09/19/ciencia/1505818015_847097.html.

5 “AI researchers greeted the announcement of the Black in AI workshop at NeurIPS, a leading machine learning conference, by questioning whether the event was necessary and arguing that it would be discriminatory.”

- diversidade nas fontes de dados utilizadas – os algoritmos usados em IA em geral são conhecidos. O que diferencia os sistemas normalmente são as bases usadas para treinamento. Se as bases são enviesadas, o sistema aprende e repete o padrão;
- diversidade entre as pessoas para quem os sistemas são desenvolvidos – o sistema reconhece rostos/vozes/linguagens diversas?

Ainda que distintos, os três aspectos estão diretamente relacionados e iniciativas para criar equipes de desenvolvimento diversas podem resultar em dados menos enviesados e em sistemas adaptados para usuários diversos. Na próxima seção, comentamos algumas iniciativas existentes para diversificar quem faz a IA.

INICIATIVAS PARA AMPLIAR A DIVERSIDADE EM IA

Diversas iniciativas surgiram para tentar incentivar a presença de grupos sub-representados na inteligência artificial. Na grande maioria, os grupos estão mais presentes na América do Norte.

- **Black in AI**⁶: começou em 2017 como um grupo no Facebook e hoje tem mais de 5 mil membros. Organiza workshops na NeurIPS todo ano. Tem iniciativas de mentoria e ajuda financeira para participar de eventos da área. O grupo também oferece ajuda para lidar com vistos e outras barreiras burocráticas;
- **Queer in AI**⁷ – organiza workshops regularmente desde 2018. Oferece bolsas de estudo para estudantes LGBTQIA+ em todos os níveis. Também oferece ajuda para inscrições em programas de pós-graduação;
- **Latinx in AI**⁸ – surgiu em 2018 e organiza diversos workshops anualmente e grandes conferências na área de IA. Oferece mentorias e financiamento para participação nas conferências principalmente para estudantes provenientes da América Latina;
- **Women in Machine Learning**⁹ – organiza workshops anuais desde 2006 junto à NeurIPS. Oferece mentoria para mulheres na área de aprendizado de máquina e mantém uma lista de mulheres que trabalham na área;
- **Women in AI**¹⁰ – começou em 2016 também como um grupo no Facebook e hoje conta com mais de 8 mil membros. Dentre as organizações listadas, é a mais voltada para a indústria. Oferece programas de aceleração de startups lideradas por mulheres, *hackathons* e programas de atualização em IA.

Alguns grupos mais recentes incluem {Dis}Ability in AI¹¹ e Indigenous in AI¹².

Nos últimos três anos, grandes conferências internacionais, como IJCAI, AAAI, AAMAS e NeurIPS, passaram a

6 Ver: <https://blackinai.github.io/>.

7 Ver: <https://www.queerinaai.com/>.

8 Ver: <https://www.latinxinaai.org/>.

9 Ver: <https://wimlworkshop.org/>.

10 Ver: <https://www.womeninaai.co/>.

11 Ver https://elesa.github.io/ability_in_AI/.

12 Ver: <https://www.indigenous-ai.net/>.

ter chamadas públicas para organização de eventos relacionados à diversidade. Além de workshops e eventos sociais devotados a grupos pouco representados, as conferências contam também com políticas de inclusão e diversidade na sua organização, levando em conta a diversidade na composição dos comitês avaliadores e nos convites a palestrantes.

Em um desses eventos, na AAAI de 2021, foi apresentado um sistema em desenvolvimento para medir o índice de diversidade de conferências (Hupont et al., 2022), levando em conta palestrantes convidados, comitês de programa e organização e autoria de artigos. A diversidade é avaliada em três dimensões: gênero (por enquanto, binário e baseado unicamente em reconhecimento de nomes), localização (país) e área de atuação (academia, indústria ou centro de pesquisa).

No Brasil ainda há poucas iniciativas específicas para o aumento da diversidade em IA. A Sociedade Brasileira de Computação (SBC) apoia desde 2007 o WIT – Women in Information Technology¹³, workshop que acontece junto ao seu congresso principal para discutir os assuntos relacionados a questões de gênero e à tecnologia da informação (TI) no Brasil. Um dos focos é debater políticas de educação e treinamento de mulheres em TI, assim como seu recrutamento em empresas e grupos de pesquisa. O Programa Meninas Digitais¹⁴ da SBC possui dezenas de iniciativas afiliadas em todas as regiões do país, sendo voltado a motivar alunas do ensino fundamental e médio a seguir carreira

na área de computação. O grupo Mulheres em IA¹⁵ foca encontros para troca de experiências entre mulheres trabalhando em IA no Brasil. A Lawgorithm¹⁶, organização voltada à pesquisa em inteligência artificial e direito, possui entre seus núcleos um dedicado a IA e gênero e outro dedicado a IA e raça. No Centro de Inteligência Artificial (C4AI)¹⁷, temos desde 2021 um comitê de diversidade e inclusão que, além de ações de conscientização, financia bolsas de estudo para estudantes de graduação da Universidade de São Paulo. Por entendermos que a diversidade é necessária em muitos aspectos diferentes, procuramos selecionar estudantes que venham de áreas de conhecimento diversas, com variabilidade de gênero e raça. Dentro das ciências exatas, por exemplo, privilegiamos a escolha de bolsistas mulheres, dada a sua pouca representatividade.

GÊNERO EM IA

A preocupação com a falta de inclusão de mulheres em IA não é nem um pouco nova nem restrita às técnicas mais modernas. Já nos anos 80, pesquisadoras se depararam com estereótipos de gênero em sistemas especialistas e também com a falta de mulheres estudando IA. Bruce e Adam (1989) já se propunham a estudar até que ponto os sistemas reproduzem questões de gênero e quais as oportunidades para que mulheres participem do desenvolvimento dessas tecnologias.

13 Ver: <https://csbc.sbc.org.br/2023/wit/>.

14 Ver: <https://meninas.sbc.org.br/>.

15 Ver: <https://linktr.ee/mulheres.em.ia>.

16 Ver: <https://www.lawgorithm.org.br/>.

17 Ver: <https://c4ai.inova.usp.br/>.

Hall e Lovegrove (1988) discutem a preocupação com a falta de representatividade das mulheres nas equipes que desenvolvem a IA desde a época da formação. Como grande parte dos profissionais vem da computação e já era perceptível na época a diminuição de estudantes mulheres na área, como evitar que a IA fosse feita apenas por homens?

Recentemente, o Observatório de Inteligência Artificial da Organização para a Cooperação e Desenvolvimento Econômico (OCDE) publicou um artigo (Caira; Russo; Aranda, 2023) defendendo que os países-membros atuem de forma a facilitar o acesso das mulheres à educação necessária para entender como os sistemas de IA estão sendo desenvolvidos, como uma maneira de evitar que a desigualdade no mercado de trabalho aumente. O artigo também chama a atenção para o fato de que a grande maioria dos CEOs das grandes empresas de tecnologia são homens, assim como a maior parte dos pesquisadores em IA.

Em um artigo recente sobre as pessoas mais importantes por trás da IA moderna, 12 homens foram mencionados¹⁸. No Brasil, de acordo com o LinkedIn, 25% das pessoas trabalhando em IA eram mulheres em 2022¹⁹.

Com o avanço da IA generativa, a internet foi inundada de *deep fakes*, conteúdo falso em formato de fotos ou vídeos, muitas vezes contendo pornografia ou imagens altamente sexualizadas de mulheres reais

ou criadas por IA. Lamensch (2023) conta que um relatório publicado em 2019 por DeepTrace indicou que, de 15 mil vídeos classificados como *deep fake*, 96% continham pornografia com rostos de mulheres colocados artificialmente nas imagens. Ainda no mesmo artigo, Lamensch descreve um experimento feito com um aplicativo de criação de avatares a partir de fotos pela repórter Melissa Heikkilä. Quando ela tentou criar um avatar de si mesma, o aplicativo propôs avatares seminus e sexualizados. Por outro lado, os avatares dos colegas homens de Heikkilä estavam totalmente vestidos e pareciam astronautas, exploradores e inventores. Segundo Lamensch, a maior parte dos algoritmos de geração de imagem a partir de texto é treinada com a base de dados LAION-5B²⁰, com mais de 5 bilhões de pares de imagem/texto extraídos da internet. Como a internet está repleta de conteúdo misógino e racista, com falta de representatividade de diversos grupos, isso é propagado para as ferramentas que utilizam a base.

Não é apenas a IA generativa que trata mulheres de forma diferente. Um estudo mostrou que os algoritmos de classificação de imagens usados nas redes sociais para evitar imagens ofensivas classificam como inadequadas imagens de mulheres em que aparecem pequenas porções de pele, enquanto imagens de homens de torso nu não são censuradas (Mauro; Schellmann, 2023). Barrigas de grávidas são classificadas como tendo alta probabilidade de conter conteúdo sexual. O mesmo acontece com campanhas para

18 Ver: <https://www.nytimes.com/2023/12/03/technology/ai-key-figures.html>.

19 Ver: https://www.linkedin.com/posts/linkedin-noticias_iwd2023-activity-7039627442877083648-Pl-F?utm_source=share&utm_medium=member_desktop.

20 Ver: <https://laion.ai/blog/laion-5b/>.

prevenção de câncer de mama. O impacto dessas classificações nas redes sociais pode ser muito maior do que imaginamos. Muitas vezes, a imagem tida como possivelmente inadequada não é bloqueada, mas sofre um processo conhecido em inglês como *shadowbanning*, em que o algoritmo restringe o alcance de uma postagem. Enquanto o bloqueio de conteúdo considerado apropriado é notificado ao usuário, no caso de *shadowbanning* isso normalmente não acontece. O artigo cita dois casos de mulheres cujo trabalho depende da divulgação em redes sociais e que sofrem constantemente com o *shadowbanning*. Assim, suas postagens não têm o mesmo alcance que as postagens de concorrentes do sexo masculino, com consequente aumento na disparidade econômica.

Outro fenômeno que atraiu atenção no final da década passada foi o surgimento de assistentes digitais baseadas em IA, como Siri e Alexa, entre outras. Não é coincidência o fato de praticamente todas as assistentes terem nomes e vozes femininas. De acordo com West, Kraut e Ei (2019), para justificar a decisão de tornar os assistentes de voz femininos, empresas como a Amazon e a Apple citaram trabalhos acadêmicos que demonstram que as pessoas preferem uma voz feminina a uma voz masculina. Por outro lado, outros estudos sugerem que a maioria das pessoas percebe as vozes femininas como cooperativas, além de úteis, enquanto as vozes masculinas são consideradas autoritárias. Ou seja: as assistentes digitais devem ter vozes femininas para os consumidores se sentirem à vontade para dar ordens a elas! Pesquisadores observaram que o uso de assistentes virtuais ensina (inclusive a crianças), que pessoas do gênero feminino devem ser subservientes e responder a comandos. Ou

seja, a tecnologia pode servir para amplificar as desigualdades de gênero (West; Kraut; Ei, 2019). Aqui no Brasil temos a BIA²¹, a IA do Bradesco, que foi apresentada desde o início (em 2018) como feminina. Após diversas interações com usuários ofensivas e agressivas, o próprio Bradesco teve de lançar um programa contra o assédio que sua assistente sofria²².

O último ponto que eu gostaria de mencionar nesta seção é a disparidade de gênero no uso de IA, notada em um artigo recente (Costa, 2023). Uma das razões levantadas é a de que, como os estudos em áreas como a computação e a engenharia são dominados por homens, as mulheres se sentem menos à vontade para adotar novas tecnologias. Outro ponto que tem aparecido na mídia é que as mulheres têm mais medo de serem substituídas por IA em seus empregos do que os homens e, portanto, veem a tecnologia com desconfiança. Esse medo não é infundado, como mostram estudos (Unesco/OECD/IDB, 2022).

O CASO DO PROCESSAMENTO DE LINGUAGEM NATURAL

Processamento de linguagem natural (PLN) é a área da inteligência artificial que lida com interpretação e geração de linguagem “humana”. O termo “natural” aparece em contraste com a ideia de linguagem formal, como a lógica matemática ou uma linguagem de programação,

21 Ver: <https://banco.bradesco/bia/>.

22 Ver: <https://banco.bradesco/aliadosbia/>.

desenhada de acordo com regras rígidas de sintaxe e semântica e com propósitos bem definidos. As linguagens naturais, como o português e o inglês, por outro lado, surgiram de forma espontânea e foram evoluindo ao longo do tempo, de acordo com mudanças na sociedade e necessidades de comunicação entre seus falantes. PLN sempre esteve em uma posição de destaque dentro da IA, já que para a popularização do uso de ferramentas de IA elas devem ser acessadas sem a necessidade de uma linguagem formal intermediária. Um exemplo de uso são as já citadas assistentes virtuais, que recebem perguntas ou comandos em linguagem natural e também formulam a resposta na mesma linguagem.

Assim como outras áreas da IA, as técnicas de PLN atuais são quase sempre baseadas em aprendizado de máquina. Como já discutido, o uso de grandes bases de dados para treinamento dessas ferramentas tem o potencial de propagar os vieses de gênero dos textos utilizados. Uma das técnicas para predição de textos, conhecida como *word embedding*, utiliza uma representação espacial de termos e textos para fazer projeções de vetores. Assim, pode-se projetar a relação homem/rei para o espaço onde está a palavra mulher e encontrar a relação mulher/rainha. Também a partir da relação França/Paris, podemos encontrar Japão/Tóquio. No entanto, nem sempre as relações encontradas são as desejadas, como alertado por Bolukbasi et al. (2016). No artigo, os autores discutem como alguns vieses estão presentes na própria língua inglesa, mas também como sistemas treinados em bases supostamente idôneas também aprendem estereótipos de gênero.

Esse foi talvez o primeiro artigo a estudar o fenômeno em profundidade e a propor

técnicas de “desenviesamento” (*debiasing*) de algoritmos. A técnica proposta, no entanto, baseia-se no fato de que, na língua inglesa, a maior parte das palavras é neutra em relação a gênero, o que não ocorre em português. Se olharmos para palavras referentes a profissões, por exemplo, temos *teacher*, que pode significar professor ou professora, *doctor*, que pode significar médico ou médica, e assim por diante. Claramente, essa diferença entre as línguas traz uma dificuldade extra para traduções, sejam elas automáticas ou realizadas por seres humanos. Como podemos traduzir “*the teacher has arrived*”? A tradução para o português deve optar por um gênero? Ou tentar utilizar termos que mantenham a indefinição da sentença em inglês? O Google Tradutor, quando consultado sobre a tradução de “*the teacher has arrived*” para português, devolve como resposta (em janeiro de 2024): “A professora chegou”. No entanto, ao mudarmos a profissão da consulta para *doctor*, a resposta passa a ser “o médico chegou”. Aqui podemos ver o resultado do aprendizado estatístico produzindo um estereótipo de gênero – mulheres como professoras, homens como médicos. A situação fica ainda pior quando traduzimos entre duas linguagens com marcação de gênero, como português e espanhol. Traduzindo “a médica e o enfermeiro chegaram” para o espanhol, retorna como resultado “*llegaram el médico y la enfermera*”, porque as traduções entre as duas línguas são feitas com uma passagem intermediária pelo inglês (português → inglês → espanhol) e, assim, perdemos os gêneros marcados na frase original. Esse fenômeno foi estudado por Prates et al. (2020) em relação a línguas como o húngaro, que têm ainda menos marcações de gênero do que o inglês.

No caso das traduções do inglês para o português, já comentamos que elas apresentam um desafio mesmo para tradutores humanos. Existe todo um movimento social preconizando a adoção de uma linguagem inclusiva, muitas vezes chamada de linguagem neutra, que possa abarcar situações em que o gênero seja desconhecido ou não binário. Em inglês, já é bastante consensual o uso dos pronomes *they/them* para essas situações. Já em português, a criação de tal linguagem inclusiva vem despertando fortes reações contrárias, e até proibições de uso²³. Isso se deve ao fato de a bandeira ter sido levantada inicialmente por populações marginalizadas, como as pessoas transgêneros ou que se identificam com alguma das muitas possibilidades de gênero existentes para além das opções binárias (masculino/feminino) e que são excluídas pelo uso da norma binária. Mesmo sem entrar na questão sociológica, vemos que a linguagem inclusiva não é apenas uma questão de discutir gêneros não conformantes, mas também de transmitir informações quando o gênero das pessoas é desconhecido. Nesse caso, os modelos de linguagem modernos baseados em aprendizado de máquina têm pouco a dizer, pois (ainda?) não existe uma linguagem inclusiva que seja consenso e muito menos grandes conjuntos de textos bilíngues em linguagem inclusiva. Em Lopes et al. (2023), fizemos um estudo preliminar utilizando a linguagem proposta por Cassiano (2019) e testamos um protótipo de tradutor de português “correto” para um português “inclusivo”.

23 Ver: <https://g1.globo.com/educacao/noticia/2023/02/13/linguagem-neutra-barrada-pelo-stf-lei-que-proibe-o-uso-de-linguagem-neutra-existe-em-3-estados-e-2-capitais.ghtml>.

Lopes et al. (2023) realizaram também experimentos com o ChatGPT, que apresentou um bom desempenho para o Prompt “Escreva as frases seguintes usando linguagem neutra: <[frases]>”, mas utilizou alguns artifícios considerados inadequados. Exemplos positivos desse uso são as seguintes reescritas: “Aqueles que terminaram podem sair → Quem terminou pode sair” e “Nós somos alunos de teatro → Nós somos estudantes de teatro”. Exemplos negativos são: “As cantoras denunciaram os assédios → As artistas denunciaram os assédios”, no qual o gênero feminino não foi neutralizado, e “Você é biólogo? → Você é biólogx?”, no qual o uso do caractere “x” como neutralizador foi aplicado, o que não é recomendado atualmente.

DISCUSSÃO

A maior parte das questões ligadas a gênero é inerente à sociedade e não exclusividade da área de inteligência artificial. No entanto, como vimos, as novas tecnologias têm o potencial de amplificar e fossilizar estas questões. Amplificar, pois o alcance da IA aumentou e as pessoas têm a tendência a imaginar a tecnologia como sendo imune a preconceitos tipicamente humanos e, portanto, não questionar. Além disso, conforme um modelo de linguagem enviesado serve de base para a implementação de outras ferramentas, os vieses vão sendo transmitidos em cadeia. Fossilizar, porque, como os modelos de linguagem são treinados com grandes volumes de dados já existentes, mesmo que a sociedade evolua, existe o risco de os preconceitos do passado serem usados para moldar o futuro.

Para mitigar os danos, podemos pensar em propostas de curto, médio e longo prazos. No curto prazo, seria muito importante exigir das empresas transparência em relação aos dados usados no treinamento das ferramentas e formular os requisitos de representatividade que tais dados devem cumprir. A médio prazo, políticas afirmati-

vas de contratação podem gerar equipes de desenvolvimento mais diversas e, portanto, mais sensíveis ao sexismo. A longo prazo, políticas públicas de educação igualitária e de atração de meninas para as áreas de tecnologia têm o potencial de fazer com que as próximas gerações não precisem se preocupar com questões de gênero.

REFERÊNCIAS

- BENJAMIN, R. *Race after technology: abolitionist tools for the New Jim Code*. Cambridge, Polity, 2019.
- BOINODIRIS, P. *The importance of diversity in AI isn't opinion, it's math*. 2024. Disponível em: <https://www.ibm.com/blog/why-we-need-diverse-multidisciplinary-coes-for-model-risk/>.
- BOLUKBASI, T. et al. "Man is to computer programmer as woman is to homemaker? Debiasing word embeddings", in *Proceedings of the 30th International Conference on Neural Information Processing Systems (NIPS'16)*.
- BOMMASANI, R. et al. *On the opportunities and risks of foundation models*. 2022. Disponível em: <https://arxiv.org/abs/2108.07258>.
- BRUCE, M.; ADAM, A. "Expert systems and women's lives: a technology assessment". *Futures*, v. 21, issue 5, 1989, pp. 480-97.
- BUOLAMWINI, J. *Gender shades: intersectional phenotypic and demographic evaluation of face datasets and gender classifiers*. Dissertação de mestrado. Massachusetts, MIT, 2017. Disponível em: <https://www.media.mit.edu/publications/full-gender-shades-thesis-17/>.
- CAIRA, C.; RUSSO, L.; ARANDA, L. *Artificially inequitable? AI and closing the gender gap*. 2023. Disponível em: <https://oecd.ai/en/wonk/closing-the-gender-gap>.

- CASSIANO, O. *Guia para "linguagem neutra" (PT-BR)*. 2019. Disponível em: <https://medium.com/guia-para-linguagem-neutra-pt-br/guia-para-linguagem-neutra-pt-br-f6d88311f92b>.
- COSTA, M. "Why are fewer women using AI than men?". *BBC*, 2023. Disponível em: <https://www.bbc.com/news/business-67217915>.
- HALL, W.; LOVEGROVE, G. "Women and AI". *AI & Society*. v. 2, 1988, pp. 270-1.
- HUPONT, I. et al. "Monitoring diversity of AI conferences: lessons learnt and future challenges in the DivinAI project". 2022. Disponível em: <https://arxiv.org/abs/2203.01657>.
- LAMENSCH, M. *Generative AI tools are perpetuating harmful gender stereotypes*. 2023. Disponível em: <https://www.cigionline.org/articles/generative-ai-tools-are-perpetuating-harmful-gender-stereotypes/>.
- LOPES, W. R. et al. "Eu, tu, ele, ela, elu, nós, vós, eles, elas, elus por um modelo de linguagem neutra". *Anais do Encontro Nacional de Inteligência Artificial e Computacional (Eniac)*, 20, 2023. Disponível em: <https://sol.sbc.org.br/index.php/eniac/article/view/2575>.
- MAURO, G.; SCHELLMANN, H. "'There is no standard': investigation finds AI algorithms objectify women's bodies". 2023. Disponível em: <https://www.theguardian.com/technology/2023/feb/08/biased-ai-algorithms-racy-women-bodies>.
- O'NEIL, C. *Weapons of math destruction*. Londres, Penguin Books, 2017.
- PRATES, M. O. R.; AVELAR, P. H.; LAMB, L. C. "Assessing gender bias in machine translation: a case study with Google Translate". *Neural Computing & Applications*, v. 32, 2020, pp. 6.363-81.
- UNESCO – Organização das Nações Unidas para a Educação, a Ciência e a Cultura; OECD – Organização para a Cooperação e Desenvolvimento Econômico; IDB – Banco Interamericano de Desenvolvimento. *The effects of AI on the working lives of women*. Paris, Unesco, 2022.
- WEST, M.; KRAUT, R.; CHEW, H. E. "The rise of gendered AI and its troubling repercussions". *I'd blush if I could: closing gender divides in digital skills through education*. Paris, Unesco, 2019.
- WEST, S. M.; WHITTAKER, M.; CRAWFORD, K. "Discriminating systems: gender, race and power in AI". *AI Now Institute*, 2019. Disponível em: <https://ainowinstitute.org/publication/discriminating-systems-gender-race-and-power-in-ai->



O futuro da pesquisa em inteligência artificial

Anna Helena Reali Costa

Fabio Gagliardi Cozman

resumo

Este artigo discute o futuro da pesquisa em inteligência artificial, argumentando que, em um futuro próximo, técnicas simbólicas continuarão a receber certa atenção, enquanto técnicas de aprendizado de máquina baseadas em dados continuarão a apresentar crescimento explosivo. Modelos obtidos por aprendizado profundo, em especial modelos de linguagem, serão aplicados em variados setores e terão consideráveis melhoras (em eficiência, em interpretação, em desempenho). Ainda no futuro próximo, o debate social sobre inteligência artificial se tornará mais concreto, exigindo esforço de pesquisa sobre legislação, impacto social e mercado de trabalho. Procurando olhar para um futuro mais distante, entendemos que combinações entre métodos simbólicos e baseados em dados receberão maior atenção, enquanto apostas, como computação quântica, poderão levar a novos saltos de qualidade.

Palavras-chave: futuro da IA; aprendizado de máquina; aprendizado profundo; IA e sociedade.

abstract

This paper examines the future of research in artificial intelligence, arguing that, in the near future, symbol-based techniques will continue to receive some attention, while machine learning techniques based on data processing will continue to grow explosively. Models produced through deep learning, in particular language models, will be applied to many sectors and will be significantly enhanced (in efficiency, in interpretability, in performance). Still in the near future, the social debate about artificial intelligence will take a more concrete form, demanding research effort related to regulation, social impact and job markets. Looking into a more distant future, we believe that mixtures between symbol-based and data-based methods will get more attention, while some bets, such as quantum computing, may take us to new performance levels.

Keywords: future of AI; machine learning; deep learning; AI and society.

Vinte anos atrás, prever os rumos da pesquisa em inteligência artificial exigia extrapolar, para um horizonte relativamente curto, um conjunto limitado de subáreas disjuntas. De forma geral, a pesquisa seguiria algumas abordagens consagradas, por exemplo, explorando representações baseadas em lógica, ou aprendizado de máquina inspirado por organismos biológicos. Previsões para um futuro distante eram mais baseadas em ficção científica do que em pesquisa acadêmica. Porém, o horizonte próximo era bem delineado, com algumas linhas de pesquisa oferecendo algum sucesso prático, outras nem tanto.

Por volta de 2007, essa realidade começou a mudar, inicialmente de forma gradual

e recentemente de forma estonteante. Devido ao sucesso prático da tecnologia de IA, cresceu o número de pesquisadores, assim como o volume de investimentos (tanto públicos quanto privados), e sobretudo o número de abordagens e de algoritmos disponíveis na literatura. A previsão do futuro, mesmo o futuro próximo, tornou-se bem mais complexa. Prognósticos sobre IA ficaram mais arriscados, sobretudo em relação a empresas e seus produtos, mas mesmo em relação à pesquisa, que normalmente tem um tempo

ANNA HELENA REALI COSTA é professora titular do Departamento de Engenharia de Computação da Escola Politécnica da USP e coordenadora do Centro de Ciência de Dados (c²d), uma parceria entre o Banco Itaú-Unibanco e a USP.

FABIO G. COZMAN é professor titular do Departamento de Engenharia Mecatrônica da Escola Politécnica da USP e diretor do Center for Artificial Intelligence (C4AI), uma parceria entre a Fapesp, a IBM e a USP.

de maturação mais longo. Neste breve artigo procuramos tatear os caminhos prováveis da IA para o futuro imediato e arriscar algumas poucas previsões para o futuro mais distante. Para tanto, agregamos nossa experiência a uma análise dos principais temas e discussões nas grandes conferências da área, tais como NeurIPS¹, IJCAI² e AAAI³ –, notando que tais conferências hoje atraem tanto atores da academia quanto das grandes empresas de tecnologia. Estas últimas participam não só com interesse em recrutamento, mas também para compartilhar suas mais recentes inovações e para absorver novos resultados da academia.

Certamente há formas diferentes de discutir o futuro da IA, em particular o futuro mais distante. Por exemplo, o recente trabalho de Grace et al. (2024) combina previsões de mais de 2.700 pesquisadores sobre a IA para os próximos 200 anos, focando sobretudo os impactos da tecnologia em setores econômicos e sociais. Aqui fazemos um recorte focado em um prazo menor, com ênfase na pesquisa sobre a tecnologia em si.

ENTENDENDO OS CAMINHOS: *NEATS E SCRUFFIES*

Para entender o futuro, é útil ter um entendimento do passado. A área de IA não é nova, tendo se iniciado por volta de 1950, mas seus objetivos sempre foram

difíceis de especificar. Definir a área é um desafio que recentemente saltou dos livros-textos acadêmicos para o campo da diplomacia internacional. Um órgão que tem se destacado nesse debate é a Unesco, que considera como sistemas de IA aqueles que “têm capacidade de processar dados e informações de uma forma que se assemelha ao comportamento inteligente e, normalmente, inclui aspectos de raciocínio, aprendizagem, percepção, previsão, planejamento ou controle” (Unesco, 2021, artigo 1.2).

Podemos prever, sem chance de errar, um ponto sobre a IA: sua definição continuará a ser discutida infinitamente na academia, na cultura popular, na legislação, na diplomacia internacional. Para nossos propósitos aqui, a lista simples oferecida pela Unesco é bastante adequada. Estamos interessados em sistemas computacionais que representam conhecimento, raciocinam, tomam decisões, aprendem com experiências e interagem com o mundo.

Ao se defrontar com a tarefa de construir um artefato que satisfaça pelo menos a alguns aspectos da definição acima, um pesquisador pode enfatizar objetivos diferentes (Russel; Norvig, 2022). Um possível objetivo é a construção de um artefato que se comporte de maneira “racional” segundo alguma definição formal de racionalidade. Por exemplo, um agente que, ao se defrontar com um problema lógico, saiba operar de forma exata os axiomas da lógica formal. Ou um agente que, ao se defrontar com uma situação de incerteza, saiba calcular probabilidades e comparar o valor esperado de várias alternativas. Um outro possível objetivo é a simulação do comportamento “humano”, ou pelo menos o

1 Ver: <https://nips.cc/>.

2 Ver: <https://www.ijcai.org/>.

3 Ver: <https://aaai.org/conference/aaai/>.

comportamento observado de seres humanos (e, às vezes, de outros animais). Para tanto, podem ser usadas técnicas inspiradas nas propriedades biológicas, ou podem ser combinadas várias técnicas para emular aquilo que é observado: o que importa é imitar, tanto quanto possível, o que vemos no principal sistema que sabemos ser inteligente – o ser humano.

Ocasionalmente pesquisadores cujo foco é a análise formal têm sido chamados de *neats*, enquanto pesquisadores que procuram pragmaticamente combinar variadas inspirações e técnicas têm sido chamados de *scruffies*⁴ (Abelson, 1981). Nas últimas décadas, em alguns momentos a abordagem *neat* predominou, enquanto em outros momentos a abordagem *scruffy* foi vitoriosa.

Por décadas, o principal foco da pesquisa em IA foi a codificação de problemas em linguagens formais, frequentemente baseadas em lógica, e a busca por soluções usando algoritmos gerais (Brachman; Levesque, 2004). Este artigo dará pouco espaço para novos resultados sobre representação formal de conhecimento, raciocínio lógico e racional, tomada de decisões com garantias de desempenho e assim por diante; entretanto, todos esses tópicos fazem hoje parte da agenda de pesquisa de IA e sem dúvida estarão presentes na comunidade de pesquisa em IA durante o futuro próximo. Em particular, os próximos anos certamente trarão melhores algoritmos de geração de planos que nos levem de um estado inicial a um

estado desejado, bem como a melhores linguagens de representação de restrições e de relações entre objetos.

Seja como for, o principal foco da IA passou, nas duas últimas décadas, para a extração de padrões a partir de enormes bases de dados, usando métodos estatísticos ou de inspiração biológica para interpretar observações, medidas, textos e imagens. Embora tanto *neats* quanto *scruffies* estejam bem representados na comunidade de pesquisa, abordagens *scruffy* que se baseiam em dados para reproduzir comportamentos desejados estão obtendo mais vitórias e maior visibilidade.

O APRENDIZADO DE MÁQUINA CONTINUARÁ CRESCENDO FORTEMENTE

A maior disponibilidade de dados na sociedade como um todo e a maior capacidade de distribuí-los e de processá-los computacionalmente levaram, ao longo das duas últimas décadas, a um extraordinário sucesso das técnicas ditas de aprendizado de máquina (Russel; Norvig, 2022).

Sem dúvida, os próximos anos continuarão testemunhando mais aplicações de sucesso de aprendizado de máquina, bem como novas técnicas voltadas a extrair padrões e regras de massas de dados.

Vale a pena diferenciar entre duas classes de métodos de aprendizado de máquina, já que essas duas classes evoluíram e evoluirão de formas diferentes. Para tanto, é preciso compreender alguma terminologia essencial sobre aprendizado de máquina, incluindo termos como aprendizado estatístico, aprendizado profundo, redes neurais e modelos de linguagem.

4 A tradução mais adequada para o português seria, talvez, “empertigados”, para *neats*, e “desgrenhados”, para *scruffies*.

O verbo “aprender” refere-se a um vasto conjunto de fenômenos. É possível aprender lendo uma única vez uma lista de instruções específicas, ou observando um agente resolver um determinado problema algumas vezes, ou coletando uma enorme quantidade de dados, extraindo padrões e reproduzindo esses padrões em ações futuras. Essa última abordagem é a essência do *aprendizado de máquina estatístico* (Hastie; Tibshirani; Friedman, 2009), do qual se extraem padrões de massas de dados tanto por métodos estatísticos consagrados, como regressão ou agrupamento, quanto por métodos relativamente novos, como *boosting* ou SVMs.

Próximo do final da década de 2000 já era claro que as soluções baseadas em aprendizado de máquina alcançavam um desempenho notável, e frequentemente surpreendente, em diversas tarefas anteriormente consideradas desafiadoras. Por exemplo, a tradução automática de textos, por décadas abordada por meio de regras linguísticas, teve enorme evolução quando sua base passou a ser meramente estatística: coletamos sentenças em uma língua e suas traduções em outra língua, e estimamos uma função a partir das sentenças coletadas, que leva as palavras da primeira língua para seus mais prováveis correspondentes na outra língua.

Esse movimento foi muito bem capturado por um artigo influente publicado em 2009, cujo título é muito apropriado: “The unreasonable effectiveness of data”⁵ (Halevy; Norvig; Pereira, 2009). Em suma, os autores desse artigo argumentam que

é viável reproduzir comportamentos aparentemente inteligentes de forma surpreendentemente efetiva simplesmente coletando grandes quantidades de dados e emulando padrões encontrados nesses dados. Note-se que isso foi discutido em 2009! Desde então, a pesquisa em IA baseada em dados só se acelerou e se expandiu.

De certa maneira, o conjunto de técnicas agrupado sob o rótulo de “aprendizado estatístico” compõe um corpo já tradicional de ferramentas que tem encontrado muito sucesso prático. O futuro próximo da IA certamente inclui um continuado e crescente sucesso do aprendizado de máquina estatístico. Pode-se prever que a qualidade de decisões automáticas, por exemplo, em diagnóstico médico, ou em detecção de anomalias em produção industrial, terá significativa melhora nos próximos anos. Haverá um farto espaço para pesquisa em novas técnicas que possam processar mais dados de forma eficiente e sustentável.

Porém, não é no aprendizado de máquina estatístico tradicional que provavelmente serão encontradas novas mudanças de paradigma na pesquisa em IA. Para entender esse ponto, façamos uma retrospectiva sobre o avanço espetacular do *aprendizado profundo*.

Técnicas de aprendizado profundo se popularizaram a partir de 2012, ano em que redes neurais do tipo convolucional venceram uma competição de detecção de objetos em imagens (Krizhevsky; Sutskever; Hinton, 2012). Tais técnicas são baseadas em redes neurais artificiais com muitas camadas, de onde vem o adjetivo “profundo” (o termo original, em inglês, é *deep learning*). Redes neurais artificiais procuram imitar alguns aspectos da orga-

5 Esse título é praticamente impossível de se traduzir com precisão: trata-se de indicar a incrível/surpreendente/espantosa(irracional?) eficácia permitida pelo uso de dados.

nização neuronal observada em organismos biológicos, sendo compostas de camadas que se interconectam, cada camada contendo um conjunto de pequenas unidades de computação simples que são chamadas de neurônios artificiais. A analogia com cérebros biológicos, embora grosseira, é explorada há décadas em IA, em alguns momentos tendo gerado enorme entusiasmo e, em outros, tendo se deparado com ceticismo. Desde o aparecimento do aprendizado profundo, esse ceticismo praticamente desvaneceu e hoje existe farta evidência empírica indicando que redes neurais com muitas camadas têm capacidade de reproduzir padrões que métodos estatísticos mais tradicionais não conseguem aproximar.

De certa forma, a vitória atual do aprendizado profundo é uma extraordinária vitória da filosofia *scruffy*. Redes neurais artificiais são sistemas bastante opacos, cuja análise formal é extremamente complexa; é praticamente impossível explicar de maneira não trivial a razão de uma rede neural artificial produzir uma determinada saída a partir de uma dada entrada.

A efervescência trazida pelo aprendizado profundo, que teve impacto avassalador entre 2012 e 2015, foi seguida de outras pequenas revoluções em rápida sucessão. Em meados da década de 2010 tornou-se comum transformar símbolos em quantidades numéricas, o que melhorou em muito o processamento de linguagem natural⁶. Outra revolução ocorreu em 2017, com o aparecimento de uma organização

particular de neurônios artificiais e camadas que amplia a capacidade de processar dados pelas redes neurais resultantes. Tais redes são os chamados *transformers* (Vaswani et al., 2017) e, em pouco tempo, apareceram os *modelos de linguagem de grande porte* baseados em *transformers*. Um modelo de linguagem recebe símbolos em uma determinada linguagem e produz probabilidades de emissão de outros símbolos (na mesma ou em outra linguagem). Por exemplo, dada a sequência “o aluno entrou na”, um modelo de linguagem produzirá probabilidades para palavras como “sala”, “escola”, “enfermaria” etc. Modelos de linguagem são usados há décadas, porém o primeiro modelo de linguagem de grande porte que fez uso de *transformers* foi o modelo Bert (Devlin et al., 2019), que apresentou um desempenho excepcional em muitas tarefas de processamento de língua natural.

Hoje existem muitos modelos de linguagem de grande porte, indicados pela sigla LLM (do termo em inglês *large language model*). O mais conhecido é provavelmente o modelo GPT, o motor que dá substância ao popular serviço ChatGPT⁷. LLMs são espantosamente proficientes em algumas tarefas, como resumir textos, corrigir a gramática e a sintaxe de um determinado texto, mudar o tom de um texto para públicos específicos e traduzir de um idioma para outro. Ao receber textos adequadamente preparados, certos LLMs conseguem até mesmo emular padrões relativamente sofisticados de

6 Usualmente um mapeamento de símbolos a números é chamado de *embedding* na literatura (Mikolov et al., 2013).

7 Ver: <https://chat.openai.com/>.

raciocínio, embora o princípio básico de seu funcionamento seja bastante simples: em resumo, emitir os símbolos mais prováveis dado o conjunto de símbolos de entrada. Essa variedade de IA opera de forma bem diferente daquilo que imaginamos ser a inteligência humana. Além disso, os LLMs também apresentam problemas. Algumas vezes, LLMs podem incluir dados incorretos declarados como fatos, em um fenômeno chamado de “alucinação” (Maynez et al., 2020). Isso ocorre quando o modelo não possui a compreensão contextual necessária sobre o que foi perguntado. Um LLM não consegue verificar se uma informação está correta ou não e pode mesmo fornecer respostas conflitantes a uma pergunta se esta for feita mais de uma vez. Pesquisadores têm explorado um conjunto de ideias para reduzir o número de alucinações: coleta de mais dados e seleção mais apurada de dados; quantificação da incerteza no processo de geração de palavras; e melhora do processo de estimação de probabilidades de palavras (Kaddour et al., 2020; Ziwei et al., 2023).

O futuro da IA passa por um melhor entendimento dos LLMs e das técnicas a eles associadas. Alguns dos problemas enfrentados por LLMs poderão ser superados com maior coleta de dados. Mas construir um LLM pode ser extremamente custoso tanto em termos computacionais quanto em energia. O significativo gasto de energia para treinamento de LLMs foi inicialmente notado por Strubell, Ganesh e McCallum (2019), um estudo que levou a muitos outros experimentos nos últimos anos. Parece seguro dizer que LLMs têm um custo ambiental, tanto para treinamento quanto para inferência

(Desislavov; Martínez-Plumed; Hernández-Orallo, 2023), mesmo quando examinados a médio prazo (Patterson et al., 2022). Além disso, o custo ambiental em outras dimensões, por exemplo, relativo ao consumo de água, deve ser mais bem equacionado no futuro (IEA, 2024).

Dado o incrível esforço que vemos hoje em torno dos desafios de superar os pontos fracos dos LLMs, podemos esperar para o futuro próximo uma nova geração de modelos, mais eficientes no uso de energia, com maiores garantias de veracidade em suas afirmações e com maior habilidade de se conectarem com programas auxiliares – por exemplo, programas que facilmente resolvam questões matemáticas. Haverá significativo esforço para desenvolver instrumentos teóricos que permitam analisar e compreender modelos baseados em redes neurais profundas e técnicas correlatas, como mecanismos de geração de imagens. Haverá também pesquisas sobre formas de organizar esses modelos de maneira modular, assim combinando competências de modelos diversos. Sobre tudo, veremos intenso esforço de pesquisa centrado em modelos de linguagem que contenham mecanismos de controle e prevenção de alucinações. Isso nos levará aos tão esperados auxiliares artificiais que nos darão apoio e aconselhamento sem absurdas divagações. Muitas empresas têm mostrado que pretendem levar esses avanços para o mercado com celeridade. Com efeito, uma previsão fácil para o futuro próximo é a chegada ao mercado de variados produtos que usam aprendizado de máquina (em particular LLMs). Esse avanço será acompanhado por um melhor entendimento de quais técnicas

são de fato úteis, transcendendo o atual período de agitação intensa.

Seja como for, entre métodos estatísticos e neurais, o futuro próximo da pesquisa em IA será em boa parte voltado a aplicações práticas. Na academia veremos cada vez mais a IA aplicada à compreensão e à antecipação de fenômenos físicos ou sociais, nos mais variados campos do saber. Exemplos simples: a IA pode combinar massas de dados e conhecimento sobre oceanos para prever ressacas e prevenir populações litorâneas; pode ainda combinar milhões de dados do SUS para encontrar a melhor estratégia de vacinação. Veremos também um continuado esforço para tornar menos opacas as decisões tomadas por métodos estatísticos ou neurais. Ou seja, esforços para facilitar a interpretação dessas decisões ou para gerar explicações para essas decisões.

O DEBATE SOBRE IA E SOCIEDADE SE TORNARÁ MAIS CONCRETO

Nos últimos anos a IA tem suscitado diversas preocupações e gerado inúmeros debates. A sociedade tem discutido o impacto da IA no jogo político, bem como examinado a capacidade da IA de produzir falsidades e de infringir a privacidade. Avanços recentes na tecnologia levantam considerações éticas e preocupações sobre o uso responsável da IA e a necessidade de evitar usos maliciosos. Modelos estatísticos e modelos de linguagem podem ser suscetíveis a ataques mal-intencionados que manipulem dados de entrada para alterar decisões. Por serem extremamente complexos, modelos obtidos via aprendi-

zado profundo são difíceis de interpretar, o que dificulta a detecção de vulnerabilidades e de vieses. Experimentos realizados nos últimos anos detectaram decisões de caráter discriminatório em sistemas de IA destinados a aplicações práticas.

Efeitos no mercado de trabalho têm também causado preocupação. O relatório sobre inteligência artificial e empregos, divulgado pela Organização para a Cooperação e Desenvolvimento Econômico, alerta que 27% dos empregos dos seus países membros são de profissões com alto risco de automatização pela IA (OECD, 2023). Recomenda-se ali que nações invistam em capacitação apropriada para trabalhadores.

O futuro próximo verá essas preocupações se transformarem em instrumentos de controle e regulamentação ao redor do mundo. O debate, que até agora se deu mais no campo das ideias, começará a produzir leis e diretrizes concretas. Em toda parte, a legislação será uma ferramenta discutida com maior intensidade, especialmente no que diz respeito à responsabilização por erros realizados pela IA e à quantificação de riscos associados à IA. A considerável discussão em torno do Artificial Intelligence Act aprovado pela União Europeia (2024), que terá efeitos mais adiante, indica o quanto o assunto causa controvérsia.

Entretanto, não é possível prever claramente quais desses esforços terão resultados positivos ou negativos. Não será fácil encontrar o balanço adequado entre incentivo da IA “boa” e controle da IA “ruim”. Os países que adotarem uma abordagem mais cautelosa e ponderada provavelmente alcançarão mais sucesso. Esses movimentos exigirão suporte acadêmico em áreas

como direito, economia e ciências sociais e políticas. Compreender como o ser humano reage a esse novo ambiente ao seu redor, que inclui máquinas com alguma inteligência, e encontrar formas de tornar esse ambiente mais humano e mais justo irão requerer pesquisa séria e continuada.

Uma previsão para o futuro próximo: em breve veremos esforços de empresas, em todos os setores, para avaliar e certificar suas tecnologias de IA como socialmente adequadas. Processos de certificação variados serão tentados: autorregulação, órgãos certificadores, órgãos centrais etc. A regulação via poder público não será o único caminho no futuro.

Outra previsão para o futuro: as nações que optarem por proibir a tecnologia enfrentarão condições econômicas e sociais mais desfavoráveis nos próximos anos do que aquelas que adotarem a IA, mas agirão com cautela.

Como estamos pensando em nações, cumpre aqui anotar um risco trazido pela IA que se superpõe a variados outros fenômenos sociais: haverá um aumento substancial na disparidade entre nações no que diz respeito à IA, e isso não significa apenas um aumento na diferença de infraestrutura computacional relevante ou robôs de serviços ou manufatura, mas sobretudo um aumento em disparidades nas dimensões sociais como educação, trabalho e renda (Alonso et al., 2020). Essas diferenças já se aplicam hoje à pesquisa em IA: enquanto algumas poucas nações estão investindo pesadamente na pesquisa em novas tecnologias, e notadamente em IA, e lideram decisivamente, muitas outras nações permanecem à mercê de avanços produzidos fora de suas fronteiras.

Deve-se notar também uma mudança significativa ocorrida ao longo dos últimos anos e que tem importantes implicações práticas. Até cerca de 2017, a maior parte dos avanços em IA acontecia na academia e em um distribuído conjunto de laboratórios empresariais com longo histórico, por exemplo, em empresas como IBM e Microsoft. Em cerca de cinco anos essa realidade mudou: hoje grande parte da pesquisa de ponta, e em particular uma enorme fração da pesquisa de maior impacto, ocorre em um conjunto de laboratórios de pesquisa de empresas, nas próprias linhas de desenvolvimento de produtos de grandes empresas, e em startups que já nascem com investimentos de porte. Algumas poucas empresas, a grande maioria baseada nos EUA, se projetaram enormemente e hoje lideram essa corrida tecnológica, com OpenAI, Microsoft e Google tendo um notável destaque. Outros países procuram avançar com rapidez, notadamente a China, mas também a Coreia, o Reino Unido e a Alemanha, para citar alguns exemplos. Tudo isso mudou a distribuição da pesquisa em IA em poucos anos, o que dificulta prever o que veremos a seguir.

Encerrando essa seção com uma pequena digressão sobre a posição do Brasil entre as nações, notamos que ele está bem integrado econômica e culturalmente com a comunidade internacional, e a população brasileira é tipicamente receptiva a novas tecnologias; nesse contexto, os avanços que impactam as nações líderes na pesquisa em IA geram impacto semelhante no Brasil. Na academia, o Brasil tem se destacado no cenário mundial de publicações e citações de artigos de IA. Em especial, em 2023, o Brasil figurou em 16º lugar em termos de publicações em IA e em 17º lugar em

termos de citações das publicações em IA, sendo o primeiro da América Latina e Caribe à frente de países como Bélgica, Suécia, Áustria e Dinamarca (OECD.AI, 2024). Têm havido esforços no Brasil para criar vários centros de pesquisa e aumentar o financiamento para inovação acadêmica e empresarial relacionada à IA. O governo brasileiro publicou em 2021 uma estratégia nacional⁸, agora em revisão, para organizar iniciativas relacionadas a apoio, financiamento e regulação. Embora o Brasil já tenha leis em vigor para abordar a proteção de dados e privacidade, há uma considerável discussão em torno de novas legislações destinadas a fortalecer o controle sobre os artefatos de IA. Apesar dos riscos que a IA acarreta, e que devem ser mitigados por uma regulamentação sensata, a tecnologia de IA pode ser um elemento significativo para impulsionar o avanço sustentável do país. Contudo, a questão mais complexa que o país enfrenta atualmente em relação à IA é como lidar com a propagação de desinformação, especialmente para fins políticos. Esse considerável número de desafios exige um investimento sério na pesquisa em IA, bem como incentivos para formação de pessoas com competência nessa área.

OLHANDO ADIANTE

Um LLM capaz de conduzir diálogos interessantes, como o GPT, é gerado a partir de bilhões de documentos coletados na internet. Supondo que colemos

muito mais documentos, mais diálogos, mais argumentos, será que chegaremos a um ponto no qual um modelo de linguagem colossal demonstre real inteligência? A médio prazo, com certeza a pesquisa em IA produzirá modelos de linguagem gigantescos – se eles de fato demonstrarão inteligência em maior profundidade é uma questão em aberto.

Mas o futuro pode nos levar a outro cenário. Pode ser que a capacidade dos LLMs evolua em muitas direções, por exemplo, em menor consumo energético, sem realmente mudar seu patamar de raciocínio. Ou seja, pode ser que as estruturas hoje empregadas não consigam atingir o desempenho que esperamos de uma real inteligência (algo em si bastante difícil de definir...). Como será o futuro da pesquisa nesse caso? Uma possibilidade é que novas estruturas consigam aprender a raciocinar a partir de textos e exemplos. Talvez seja necessário combinar modelos, reproduzindo estruturas sociais, de forma a gerar consensos e resolver problemas. Outra possibilidade é que seja necessário revisitar o uso de conhecimento e raciocínio formal, por exemplo, usando regras lógicas explícitas, para dar o próximo grande salto de desempenho em IA. Vários caminhos se abrem.

Um futuro possível é aquele no qual a IA *neat* reassumirá uma posição de destaque na construção de agentes artificiais inteligentes. Nesse caso, LLMs e seus companheiros passam a prestar serviços de alta qualidade, porém muito especializados. Sistemas práticos poderão usar diferentes técnicas, algumas neurais, outras simbólicas. Outro futuro possível é aquele no qual a conexão entre redes neurais e técnicas simbólicas ocorrerá

8 Ver: https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/arquivos/inteligenciaartificial/ebia-documento_referencia_4-979_2021.pdf.

através de uma simbiose mais complexa, tornando real o que hoje se pesquisa sob o rótulo de “IA neurossimbólica” (D’Avila Garcez; Lamb, 2023).

Seja como for, a IA não é feita só de imitações do comportamento humano; também esperamos frequentemente que aparatos artificiais tenham melhor desempenho que nós próprios em tarefas específicas. Esperamos que um automóvel corra mais do que um ser humano; esperamos que uma inteligência artificial consiga realizar inferências lógicas de forma mais confiável que um ser humano. Nesse sentido, haverá sempre espaço para técnicas simbólicas e abordagens *neats* dentro da IA. O futuro mostrará se esse espaço será pequeno e escondido, ou se ocupará a maior parte do campo.

Uma previsão fácil é que todos esses caminhos serão trilhados por pesquisadores da área de IA. Parece claro que uma união entre as abordagens *neat* e *scruffy* pertencerá ao futuro da IA.

Como a área de IA tem atraído muita atenção e muitos investimentos, mudanças súbitas devem ser esperadas a médio prazo. A previsão é arriscada; infelizmente, ainda não conhecemos nenhuma inteligência que consiga de fato prever o futuro.

Com o tempo, a pesquisa em IA procurará desenvolver auxiliares artificiais ubíquos, capazes de conduzir conversas longas e tomar decisões complexas. Essa IA se tornará mais integrada ao nosso ambiente, mais econômica e sustentável; espera-se que a pesquisa em regulação, ética, trabalho e temas ligados à sociedade também contribua para que cheguemos a uma IA positiva para o ser humano. Ou seja, olhando mais adiante, só se pode especular. Há no momento esperança de que uma união entre computação quântica e IA possa levar a um novo salto de desempenho nesta última, mas a pesquisa é incipiente nessa direção. Outra aposta de longo prazo é o desenvolvimento de conexões diretas entre máquinas e cérebros biológicos, tanto para permitir que pessoas com deficiências específicas possam recuperar certas habilidades, como para expandir a capacidade cognitiva humana. Previsões ainda mais ambiciosas que, por exemplo, se aventuram sobre máquinas superinteligentes que podem causar catástrofes, ou que imaginam artefatos com consciência, estão no momento mais próximas da ficção científica do que da ciência e tecnologia da IA.

REFERÊNCIAS

- ABELSON, R. P. "Constraint, construal, and cognitive science". *Third Annual Conference of the Cognitive Science Society*, 1981, pp. 1-9.
- ALONSO, C. et al. *Will the AI revolution cause a great divergence?* 11/set./2020. Disponível em: <https://www.imf.org/en/Publications/WP/Issues/2020/09/11/Will-the-AI-Revolution-Cause-a-Great-Divergence-49734>.
- BRACHMAN, R. J.; LEVESQUE, H. J. *Knowledge representation and learning*. San Francisco, Morgan Kaufmann, 2004.
- COZMAN, F. G. "No canal da inteligência artificial – Nova temporada de desgrenhados e empertigados". *Estudos Avançados*, v. 35, n. 101, 2021, pp. 7-20.
- D'AVILA GARCEZ, A.; LAMB, L. "Neurosymbolic AI: the 3rd wave". *Artificial Intelligence Review*, 2023.
- DESISLAVOV, R.; MARTÍNEZ-PLUMED, F.; HERNÁNDEZ-ORALLO, J. "Trends in AI inference energy consumption: beyond the performance-vs-parameter laws of deep learning". *Sustainable Computing: Informatics and Systems*, v. 38, 2023. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2210537923000124>.
- DEVLIN, J. et al. "Bert: pre-training of deep bidirectional transformers for language understanding". *Proceedings of NAACL-HLT*, 2019, pp. 4.171-86.
- GRACE, K. et al. "Thousands of AI authors on the future of AI". arXiv 2401.02843v1 (cs.CY), 2024.
- HALEVY, A.; NORVIG, P.; PEREIRA, F. "The unreasonable effectiveness of data". *IEEE Intelligent Systems*, v. 24, n. 2, 2009, pp. 8-12.
- HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The elements of statistical learning*. 2ª ed. Nova York, Springer, 2009.
- IEA – International Energy Agency. *Electricity 2024: analysis and forecast to 2026*. Disponível em: iea.org/reports/electricity-2024.
- KADDOUR, J. et al. "Challenges and applications of large language models". arXiv 2307.10169 (cs.CL), 2023.
- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. "ImageNet classification with deep convolutional neural networks". *Advances in Neural Information Processing Systems*, v. 25, 2012.
- MAYNEZ, J. et al. "On faithfulness and factuality in abstractive summarization". *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 1.906-19.
- MIKOLOV, T. et al. "Distributed representations of words and phrases and their compositionality". *Advances in Neural Information Processing Systems*, v. 26, 2013.
- OECD – Organização para a Cooperação e Desenvolvimento Econômico. *OECD employment outlook 2023: artificial intelligence and the labour market*. Paris, OECD Publishing, 2023.
- OECD.AI Policy Observatory. *Visualisations powered by JSI using data from OpenAlex*. Disponível em: www.oecd.ai.
- PATTERSON, D. et al. "The carbon footprint of machine learning training will plateau, then shrink". *Computer*, v. 55, n. 7, jul./2022, pp. 18-28.

- RUSSELL, S. J.; NORVIG, P. *Inteligência artificial – Uma abordagem moderna*. 4ª ed. São Paulo, Ed. GEN LTC, 2022.
- STRUBELL, E.; GANESH, A.; McCALLUM, A. "Energy and policy considerations for deep learning NLP". *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019, pp. 3.645-50.
- UE – União Europeia. *Regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts*. 21/jan./2024. Disponível em: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206>.
- UNESCO – Organização das Nações Unidas para a Educação, a Ciência e a Cultura. *Recomendação sobre a ética da inteligência artificial*. Código do documento: SHS/BIO/PI/2021/1, 2021. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000381137>.
- VASWANI, A. et al. "Attention is all you need". *Advances in Neural Information Processing Systems*, v. 30, 2017.
- ZIWEI, Ji et al. "Survey of hallucination in natural language generation". *ACM Computing Surveys*, v. 55, n. 12, 2023, pp. 1-38.

textos

Sociedade da mensagem para reconfigurar a des(informação)

Magaly Prado

E

m reiterado descontentamento ao ver os profissionais da imprensa serem taxados de produtores de *fake news* – até por conta de o termo *news* significar notícias – e a expressão gerar confusão ao ser disseminada como “notícias falsas”, o objetivo deste ensaio foi buscar uma alternativa viável para descrever o fenômeno da desinformação digital, levando em conta que a expressão “informações falsas” também não é adequada, pois, obviamente, informar não tem o intuito de desinformar. Portanto, trata-se de pôr à prova os preceitos de noticiabilidade contra toda e qualquer falta de ética espalhada nas redes.

MAGALY PRADO é jornalista da Rádio USP e fez estágio pós-doutoral no Departamento de Informação e Cultura da Escola de Comunicações e Artes (ECA) da USP.

A intenção é discutir, no campo da teoria da informação, as reflexões de Rafael Capurro (1995) sobre a teoria da mensagem ou angelética para alçar o termo “mensagem”, determinando seu significado, e trazer como proposta interpretativa seu uso no lugar de “informações falsas”, quando a referência recair em contexto da desinformação. A hipótese é trazer à tona a visão capurriana renovada de “mensagem” para a discussão de “informação” e, assim, fazer um paralelo a um debate em defesa de uma causa determinante quando se trata de resgatar o sentido de “notícias” e evitar sua identificação com *fake news*. Essas distinções são relevantes em razão dos impactos provocados pela abertura da internet – especialmente dos espaços possíveis das caixas de comentários; dos *templates* facilitados de criação na blogosfera e, principalmente, da entrada maciça nas redes sociais – à voz de qualquer internauta, que acarretou como efeito colateral, paradoxalmente, a saturação de informações e, com ela, o desenvolvimento de um ecossistema de desinformação desenfreada.

Uma das definições mais frequentemente usadas de ciência da informação, apontadas por Capurro e Hjörland (2007, p. 186) é a de Griffith (1980), em que a ciência da informação se ocupa com a geração, coleta, organização, interpretação, armazenamento, recuperação, disseminação, transformação e uso da informação. Curiosamente, são funções e atividades que, infelizmente, também podem ser atribuídas a fabricações e fomento de mensagens falsas. Como superar esse desafio em mitigar a desordem informacional é a pergunta crucial. Para tanto, vejamos um breviário da linha do tempo da informação quando é possível confrontar a disfunção informacional.

Ao transcender a teoria matemática da informação, a abordagem evolutiva da informação dentro das ciências naturais e sociais tem sido discutida em conferências internacionais. A ciência da informação (CI) é vista, nesse contexto, como interdisciplinar ou multidisciplinar. Capurro e Hjörland (2007, p. 166) citam Marijuan (1996, p. 91): “Como uma ciência vertical putativa, cria o próprio leque de subdisciplinas na superposição com outras ciências existentes: física da informação, química da informação (computação molecular), bioinformação (vida artificial), neurociência informacional (inteligência artificial)”. Os autores acreditam que as discussões sobre o conceito de informação em outras disciplinas são muito importantes para a CI porque muitas teorias e abordagens em CI têm suas origens em outras áreas. O conceito epistemológico de informação “põe em jogo processos de informação não humanos, particularmente na física e na biologia. E, vice-versa, os processos psíquicos e sociológicos de seleção e interpretação devem ser considerados usando-se parâmetros objetivos”, ou seja, “deixando de lado a dimensão semântica, ou, mais precisamente, considerando-se parâmetros objetivos ou situacionais de interpretação” (Capurro; Hjörland, 2007, p. 150).

O PAPEL DAS TEORIAS CRÍTICAS

Entretanto, foi justamente o abandono das questões semânticas que incitou as críticas mais agudas à CI, por parte das chamadas teorias críticas, em cujo campo a perspectiva marxista é precisamente aquela que mais se consolida (Araújo, 2009, p. 5). Na visão do autor, os “modelos anteriores, principalmente

o sistêmico, de natureza biológica, enfatizavam a estabilidade, a permanência (por meio da definição de leis, do estabelecimento das funções) e a integração (cada parte exercendo seu papel para a manutenção do todo)”. Na direção oposta, a teoria crítica enfatiza “o conflito, a desigualdade, o embate de interesses em torno da questão da informação – e, para tanto, buscará explicar os fenômenos a partir de sua historicidade” (Araújo, 2009, p. 5).

Assim, o estudo da informação, do ponto de vista crítico, não fica mais atrelado “às condições de eficácia de seu transporte, de suas funções para o equilíbrio social ou dos procedimentos funcionais para seu processamento no âmbito dos sistemas” (Araújo, 2009, p. 196). A teoria crítica, conforme o autor, relaciona-se com a ideia de “suspeição de que a realidade tenha fundamento nela mesma. Ao contrário das aproximações ‘positivas’ ao real, a teoria crítica tem por atitude epistemológica a desconfiança, a negação do evidente, a busca do que pode estar escondido ou camuflado” (Araújo, 2009, p. 196). Uma vez que a informação é entendida, pela teoria crítica, “como recurso fundamental para a condição humana no mundo, a primeira percepção que ela traz é sobre a desigual distribuição entre os atores sociais” (Araújo, 2009, p. 197). Dessa forma, nada foge aos olhos envolvendo “a questão da democratização da informação, do acesso à informação por parte de grupos e classes excluídos e marginalizados” e, ainda, “a criação de formas e sistemas alternativos de informação, e mesmo estudos sobre a contrainformação, como forma de rejeição aos regimes informacionais hegemônicos” (Araújo, 2009, p. 197).

Do mesmo modo, em um universo informativo, a teoria matemática do caos demonstra que, por um lado, os modelos nunca podem contemplar todas as variantes dos elementos que interferem em um ato comunicativo, porque tudo está inter-relacionado. “Isso seria impossível, exceto para um ser onisciente. Por outro lado, a mesma teoria demonstra que os modelos não conseguem dar conta de todas as interações estabelecidas entre todos os elementos que interferem no processo de comunicação”, esclarece Sousa (2006, p. 77).

É nesse contexto que o trabalho de Capurro, na área de epistemologia e ética da CI, reveste-se de prestígio sociocultural por exaltar a necessidade do fator interpretação da mensagem em processos de comunicação informacional, uma vez que se responde sempre a uma mensagem, sendo ela um sinal com significado específico, independente de uma eventual solicitação. Na percepção de Capurro e Hjörland (2007, p. 193), a distinção mais importante capaz de levar a novos patamares é aquela entre “informação como um objeto ou coisa (por exemplo, número de bits) e informação como um conceito subjetivo, informação como signo; isto é, como dependente da interpretação de um agente cognitivo”. Os autores pontuam ainda que a “visão interpretativa desloca a atenção dos atributos das coisas para os mecanismos de liberação para os quais aqueles atributos são relevantes” (Capurro; Hjörland, 2007, p. 193).

“A partilha de informação necessita de um suporte comunicacional para se efetivar. Isto é, a informação depende da comunicação. Não há informação sem comunicação”, assegura Sousa (2006, p. 24). Ainda no sentido de partilhar, para Gill e Adams (1998 apud Sousa, 2006), “a comunicação não é

apenas uma troca de informações ‘duras’, mas também a partilha de pensamentos, sentimentos, opiniões e experiências”. Os autores lembram que a maneira com que nos desenvolvemos como indivíduos sempre dependeu do grau de eficácia na construção de redes. A comunicação liga-nos à rede de seres humanos, começando na família e prosseguindo pelos amigos, com a ajuda da mídia, e pela sociedade.

DA HERMENÊUTICA À ANGELÉTICA

Recorrer à hermenêutica como forma de ampliar as possibilidades teóricas do estudo da informação é o que Capurro faz ao reconhecer “a existência de três paradigmas no campo: o primeiro, que denomina paradigma físico; o segundo, que identifica como paradigma cognitivo; e o terceiro, ao qual ele próprio se filia, denominado paradigma social” (Capurro, 2003 apud Araújo, 2009, p. 201).

“Criticando os dois primeiros, o autor subverte a ideia comum que se tem de que a informação é algo prévio que cria o conhecimento, propondo que, na verdade, o que ocorre é o contrário, pois a informação é o conhecimento em ação – informação é contextualizar o conhecimento, algo só pode ser considerado informação pelos atores sociais se já se tem conhecimento daquilo que é definido como informação” (Araújo, 2009, p. 201).

Pelo prisma de Capurro, a informação não seria, então, “o produto último de um processo de representação, nem algo que é transportado de uma pessoa a outra ou

mesmo algo específico para uma subjetividade isolada”. Antes de tudo, ela seria “uma dimensão fundamental da forma como os seres humanos compartilham o mundo com os outros” (Capurro, 1992 apud Araújo, 2009, pp. 201-2). É bom frisar que, na cultura da partilha digital, retuitar, replicar, compartilhar, entre outras, são formas de endosso.

Para Capurro (2010), toda hermenêutica pressupõe uma angelética. A palavra “angelética” é derivada do grego *angelía*, que significa mensagem. “A angelética designa, diferentemente da angelologia, o estudo do fenômeno das mensagens e mensageiros, independentemente de sua origem divina, ou melhor, estuda esse fenômeno dentro dos limites da condição humana”. O autor continua:

“Isso não significa que estudos relacionados sejam excluídos tanto no campo das religiões quanto nas ciências naturais (por exemplo, o estudo das mensagens e mensageiros no nível celular), mas o foco específico é o das mensagens e mensageiros humanos. O humano sempre implica naturalmente a técnica. É justamente neste momento histórico do surgimento da internet que o fenômeno técnico das mensagens e mensageiros digitais, começando com o e-mail e terminando com todo tipo de atividade política, científica, cultural e econômica na internet, desempenha um papel preponderante e, por assim dizer, paradigmático na sociedade mundial do século XXI” (Capurro, 2010).

Naturalmente, “é possível investigar o dinamismo da vida e dos seres vivos entre si e com seu ambiente como processo de elaboração de mensagens ou informações” (Capurro, 2010). Hoje, com mais alcance

e velocidade, abrimos um parêntese para as desinformações. “Isso foi amplamente demonstrado pelo desenvolvimento da cibernética [Wiener, 1948] e da teoria dos sistemas” (Capurro, 2010). Mas também é evidente que o fenômeno das mensagens “no nível humano nos confronta com uma complexidade e uma especificidade que não podem ser tratadas adequadamente apenas com base nas ferramentas conceituais e técnicas” (Capurro, 2010).

Antes de tudo, ressaltamos que tanto a hermenêutica quanto a angelética estão implicadas no significado de “mensageiro” na mitologia que as envolve (conforme o *Dicionário Houaiss*, Angelia era a Daemon que personificava as mensagens, as notícias e as proclamações). Toda interpretação pressupõe um processo de transmissão de mensagens. “O texto a ser interpretado deve ser previamente transmitido e anunciado. Hermes é principalmente um mensageiro e, nessa base, também um intérprete e tradutor” (Capurro, 2010). Esse caráter mensageiro da comunicação e da interpretação é justamente o que a angelética quer analisar, “tarefa não menos complexa e abrangente que a da hermenêutica do século passado. A angelética, diferentemente da hermenêutica, tem um caráter eminentemente prático. Não se trata apenas de entender, mas de causar uma mudança no recipiente” (Capurro, 2010).

“A relação entre remetente e destinatário pode ser concebida em analogia com o círculo hermenêutico como círculo angélico. Todo receptor é um remetente em potencial e, portanto, também um mensageiro e vice-versa. É também por isso que a dimensão ética é mais imediata na angelética do que o que Hans-Georg Gadamer pensava a respei-

to da hermenêutica com relação ao ‘círculo hermenêutico’ explicitado por ele com base na analítica heideggeriana” (Capurro, 2010).

A angelética “como teoria das mensagens é, portanto, em si uma mensagem com aspirações de criar sobre ela e sobre os outros um conhecimento comum. Suas questões dizem respeito à origem, propósito e conteúdo das mensagens” (Capurro, 2010), bem como a “estruturas de poder, técnicas e meios de disseminação, modos de vida, história(s) de mensagens e mensageiros, codificação e interpretação, aspectos sociológicos, psicológicos, políticos, econômicos, estéticos, éticos e religiosos” (Capurro, 2010).

Para o autor, que passa da filosofia da informação à filosofia da mensagem, a angelética “opera na diferença de emissor/receptor com base na crença de entendimento, ou, mais genericamente, de um processo de seleção entre dois sistemas possíveis”. A hermenêutica, inclusive a digital, “opera entre a diferença do pré-entendimento e a interpretação baseada na crença de que aquilo que é objeto no processo de interpretação tem sido transmitido com sucesso, ou seja, se ofereceu para o receptor como um objeto de seleção” (Capurro, 2010).

Apontar para os princípios e ideais da hermenêutica e, ademais, para os da angelética tem hoje uma função estratégica, pois apontam pelo avesso e, por consequência, comparativamente demonstram o quanto esses princípios estão em falta nos ambientes das redes sociais, em que as pessoas compartilham mensagens falsas porque não verificam seu conteúdo e retroalimentam a mentirinha. São nesses ambientes que pode ser testemunhado a que nível de deterioração a lógica comunicacional é levada, produzindo

impactos nocivos na vida social, especialmente na formação da cidadania necessária ao exercício da democracia, conforme será discutido a seguir.

INSIGHTS DO DESENVOLVIMENTO CÉLERE DAS REDES PRÉ E PÓS-DATAFICAÇÃO

A cibercultura exacerba o desenvolvimento da hipermídia, a linguagem própria das redes – um espaço de informação, de práticas, de interações e de processos gerados e articulados em rede – e, em tempos de ocupação das redes sociais digitais, faz aflorar o contexto de hipermobilidade.

Mesmo sabendo que a produção em rede traz ainda um rastro forte dos modos analógicos, cabe ressaltar que, no ambiente do ciberespaço (Gibson, 1984; 2008), a prática comunicacional precisa ser pensada digitalmente, considerando a emergência da multiplicidade de ferramentas, bem como *affordances* para a produção de mensagens, outros procedimentos recentes de captura e apresentação de dados e novos modelos de pensamento e sociabilidade – que, por sua vez, produzem demandas variadas em termos de audiência e sua relação com os conteúdos de informação.

A primeira fase – web 1.0 – é a da publicação em portais, sites, canais, fóruns, chats, além de mensagens por e-mail; produção de álbuns de fotos; primeiros sistemas de busca e sistemas de criptografia, entre outros. A segunda fase – web 2.0 – é a da cooperação, com redes de relacionamento, emoticons, blogs, transferência de arquivos, marketing viral, escrita coletiva, fugacidade

e (pós) convergência. Na cibercultura, surge o espaço ilimitado (como é a rede no todo) para as produções que, antes, não tinham visibilidade suficiente para uma divulgação maior. São elas as produções independentes, os ativismos artísticos e mesmo a própria possibilidade de formar redes de cidadãos conectados (Prado, 2012).

É preciso esclarecer que o intuito deste estudo não é realçar as transformações dessas primeiras fases da comunicação digital e a transposição (digitalização) de conteúdo impresso para a web; nem os posteriores conteúdos originais, porém calcados no modelo do impresso, o proeminente hipertexto e a adoção sistemática dos hiperlinks a convidar para a navegação; nem a pós-convergência que abarca o audiovisual, a qual, além de texto e fotos, traz imagens não somente em movimento, como também animadas e mutantes, conforme atualização de dados tanto por profissionais quanto por colaboradores participativos em *tagueamentos* da *folksonomia* e da taxonomia, e a entrada da multimídia; nem quando traz não somente a inserção e mineração das bases de dados nas práticas comunicativas, como também suas decorrências. A ênfase é dada à fase da comunicação móvel, à era da aceleração e da maior concorrência na circulação de informações e notícias produzidas e consumidas nos dispositivos colados aos corpos (e sua portabilidade, sensibilidade ao toque, diferentes tamanhos de telas [pequenas]) e ainda com especial destaque às ferramentas geolocais (possibilitando o rastreamento para melhor atuação dos algoritmos), tamanha é a força de sua penetração no dia a dia das pessoas, dentro dos sistemas circulatórios do tráfego na internet, na ansiedade de informação.

Conceito muito abordado desde os anos 1990, a multimídia também pode ser definida seja pelo viés da comunicação, seja pela informática, ou, ainda, por outros campos do conhecimento. Isso porque o termo pode representar várias situações muito comuns na era das mídias digitais. Por exemplo, para a informática, “multimídia” refere-se à possibilidade de um sistema de reunir, em uma mesma mensagem, diferentes tipos de representação de informação, como textos, imagens, áudios e vídeos. Essa reunião só é possível graças à complexidade dos computadores contemporâneos, que são capazes de transformar dados (ou código binário) em áreas clicáveis, imagens animadas etc. Todavia, na comunicação, o conceito representa uma verdadeira revolução no modo como meio e mensagem interagem em um ambiente digital. Por conseguinte, a evolução fez aflorar as *deepfakes*: mídia fabricada e produzida com inteligência artificial (IA) como o auge das mensagens falsas.

MUDS – MUNDOS DE FANTASIA

No aprofundamento do estudo das redes, o ativista Howard Rheingold (no capítulo “Desinformocracia”, de sua obra *Comunidade virtual*) aponta a diferença entre paradigma de difusão (imprensa, rádio, TV) e paradigma de rede (Prado, 2009). Rheingold chama a atenção para a íntima relação entre essa teia de comunicações livres, informais, pessoais e os fundamentos da sociedade democrática: “Em virtude da sua influência potencial nas convicções e percepções de um número tão grande de indivíduos, o futuro da Rede está ligado ao futuro da comunidade, da

democracia, da educação, da ciência e da vida intelectual – algumas das instituições humanas mais prezadas, independente da importância dada ao futuro da tecnologia da informática” (Rheingold, 1994).

Em meados da década de 1990, Rheingold definiu as comunidades virtuais como agregados sociais que emergem na web quando pessoas desenvolvem discussões abertas por determinado tempo, com sentimento humano suficiente para formar redes de relacionamentos pessoais no ciberespaço, e argumentou que no futuro as tecnologias pessoais tornariam possível aos indivíduos publicar notícias instantaneamente na web, à medida que os eventos se desenrolassem diante de seus olhos. Na época, o pensador não previu a emergência do tumulto informacional.

Por outro lado, Rheingold (2005, p. 443) mostra o lado selvagem do ciberespaço, em que dezenas de milhares de outras pessoas infelizes com a vida real constroem mundos de fantasia na internet. O surgimento das *muds* e das identidades alteradas nas comunidades virtuais foram os sinais evidentes desse descontentamento por parte dos internautas.

As *muds* eram mundos imaginários na internet em que os usuários começaram a viver virtualmente em comunidades, agregados sociais nos quais famílias inteiras entravam para se distrair da realidade de seus mundinhos e construir outras relações sociais. Eram comunidades de faz de conta do começo do ciberespaço, disponíveis a todos. Imagine lugares gélidos e sem sol em boa parte do ano, tendo a possibilidade de ver o pôr do sol quantas vezes quiser! Um *revival* no que hoje estão trazendo à tona o metaverso (Prado, 2022).

DIVERSAS FORMAS DE SE COMUNICAR PROLIFERAM EM CONSTANTE RENOVAÇÃO

Por consequência, o problema também recai, de um lado, na utilização de ambientes como as plataformas de redes sociais, de maneira consistente e consciente para acompanhar o fluxo apressado das mudanças, bem como a explosão de informações em modo contínuo, e, de outro lado, na dificuldade premente de estruturar uma produção atualizada e inédita, para tentar descobrir como atingir os internautas no cotidiano (ideal de todos), lembrando que eles não apenas são replicadores de informações, como também interferem na produção e customizam o que escolhem para consumir.

Mas de nada adianta tentar entender a circulação das informações ou das mensagens percorridas se a averiguação do seu êxito não for avaliada e executada. É preciso experimentar, testar e produzir para saber como funciona o que desejamos analisar.

Entendemos que é preciso sempre ver o outro lado, pôr em contraponto; e, ao concordar com o raciocínio de Santaella, o seguimos: “Tenho acalentado a ideia de que chamar atenção para o potencial construtivo das tecnologias é mais produtivo do que demonizar seus malefícios” (Santaella, 2013, p. 22). Isso exemplifica ainda de onde surgem essas percepções negativas: “[...] advêm das ruínas edulcoradas e dissimuladas do capitalismo, ruínas que Benjamin já havia vislumbrado no *Angelus Novus*, de Klee. Não é que os malefícios não existam”, e arremata airoso trazendo seu filósofo preferido, Charles Sanders Peirce, ao dizer: “Entretanto, aprendi com Peirce que

os ideais, que cultivamos em nossa mente, acabam conduzindo nosso modo de agir” (Santaella, 2013, p. 22).

Diagnosticar o estado da arte daria outro estudo, tão pertinente quanto análogo à própria rede e seu espírito – enquanto ainda era livre, com, ao menos, as pessoas participando mais e se sentindo mais inseridas no mundo. Toda a logística que interfere no cotidiano das pessoas em rede, não somente no uso das ferramentas multimídias numéricas, mas de tudo o que envolve o ato de informar dentro delas, ou seja, a linguagem, a técnica de produção, a forma de apuração dos fatos no movimento da ambulância, a navegação em camadas (de acordo com a navegação do internauta, que escolhe seus caminhos, há a possibilidade de aprofundar o conteúdo conforme a necessidade de quem navega), os aplicativos, a web semântica¹ adotada, a edição (on-line [mais rápida] ou remota), o tagueamento dentro do tagueamento etc. e, mesmo assim, sabendo que uma questão urgente de *hard news* da atualidade precisa ser resolvida e publicada hoje, de preferência, imediatamente.

Porém, a interação que nasce nas redes de computadores – cumpre assinalar – traz cooperação e conflitos, como apontado em outras reflexões (Prado, 2012). Discussões de propriedade intelectual, de um lado, liberação de conteúdo por autores em licenças específicas, de outro. Cresce a disputa entre o *copyleft* versus o *copyright* e surgem as licenças de autorias alternativas

1 O termo “web semântica” refere-se à visão do W3C da web dos dados conectados. A web semântica dá às pessoas a capacidade de criar repositórios de dados na web, construir vocabulários e escrever regras para interoperarem com esses dados.

às existentes, como as Creative Commons. Outra questão que é levantada, e que vem do movimento *cyberpunk*, é a da ética *hacker*, com os lemas “toda informação deve ser livre” e “faça você mesmo”, afluindo o ativismo e a resistência na rede (Prado, 2012). No ciberespaço também aparecem o ciborgue – homem-máquina – e a ciberarte, que propõe, entre outras bandeiras, a obra aberta e a autoria coletiva. Tudo isso influi positivamente nos internautas mais sensíveis desde o século passado. Já os impasses que a cibercultura traz à tona são muitos, como o da dominação e o do controle, da exploração e vigilância, das mensagens duvidosas, do isolamento e da sobrecarga de informações que nem sempre geram conhecimento.

A ECOLOGIA DA DESINFORMAÇÃO NA INFOSFERA E AS BOLHAS ALGORÍTMICAS

A ascensão das tecnologias digitais na infosfera se dá na proporção do crescimento dos dados em circulação. São dados produzidos por empresas, instituições e dispositivos técnicos. Contudo, entre eles, os mais graves são os dados pessoais.

Passamos por um panorama midiático no qual diversas técnicas foram (e estão sendo) adotadas na produção de informações e notícias. Dez anos atrás, Anderson, Bell e Shirky, para dar conta dessas recentes experiências de produção e consumo de informações, relatavam o uso de “análises algorítmicas de base de dados, visualização de dados, solicitações de conteúdos por parte de amadores, produção automatizada de narrativas, criação de narrativas baseadas em dados, entre outros” (Anderson; Bell; Shirky, 2013).

A proliferação do hiperfluxo das informações possui duas linhas de pensamento: uma que dá acesso a muito mais pessoas, e a outra, que muitas dessas pessoas não sabem como filtrá-las. Os algoritmos – atrapalhando e mesmo atravancando a comunicação do que realmente importa na medida em que dissipam o conteúdo no momento de sua publicação – manipulam automaticamente “mais do mesmo”, quando poderiam deixar a entrada das informações mais livre.

Na cibercultura compartilhada, as pessoas conectadas em rede escolhem quem seguir, reagem favoritando ou banindo comportamentos e produtos, personalizam também o que querem consumir. Por corolário, reforçam posicionamentos e opiniões. As escolhas feitas pelos internautas, por sua vez, geram dados e metadados que informam sucessivamente os algoritmos e os ajudam a formar as bolhas em que se inserem as pessoas (Prado, 2022, p. 34).

Além do Google e do Facebook, o aumento do número de plataformas digitais (incluindo as de mensageria instantânea, formando redes sociais em seus grupos), que facilitam a conformação de redes sociotécnicas, como Twitter, Instagram, TikTok e WhatsApp, apenas para citar algumas das mais populares no Brasil, tende a formar *clusters*, ou “bolhas” de atores, fenômeno notado por Lincoln Dahlberg (2007, apud Barcellos et al., 2017), para quem, “mesmo que o ciberespaço ofereça infinitas possibilidades, a tendência dos participantes é ir atrás de reforço às suas ideias e rejeitar as diferentes, buscar a semelhança e não a diferença”. Nesse aspecto, Dahlberg ainda afirma que os atores dispõem de mecanismos de filtro cada vez mais eficientes, cujos algoritmos remetem

a resultados de acordo com quem realizou a busca ou induzem a relacionamentos com assemelhados (Prado, 2022, p. 175).

Quando se dá a emergência do distúrbio da informação, sobretudo com o planejamento profissional em relação ao conteúdo difamatório – momento certo de ser produzido e exatamente a quem destinar –, a base de comunidades e bolhas já está preparada em escala. Basta cancelar o processo em que as premissas são imputadas nos algoritmos de IA delineados para uso da informação como estratégia de poder, escondido detrás do que é denominado “melhoria da nossa experiência”, isto é, a experiência do público (Prado, 2022, p. 200).

PROCESSAMENTO DO *BIG DATA*

O que se pretende com esta sucinta linha do tempo é levantar alguns pontos de relevo para demonstrar o quanto a sociedade da mensagem, pensada por Capurro e outros, foi gradativamente se dissipando nas redes sociais digitais. Marco decisivo dessa dissipação encontra-se no escândalo da Cambridge Analytica², exposto pelas jornalistas Carole Cadwalladr e Emma Graham-Harrison (2018), um caso de violações de regras e abuso de dados em massa para “construir um sistema que pudesse traçar o perfil individual dos eleitores americanos, a fim de direcioná-los com anúncios políticos personalizados”. Fez a empresa entrar em colapso, após o *Observer* ter revelado que se

apropriara para fins comerciais de cerca de 87 milhões de perfis no Facebook (multado posteriormente em bilhões) e dos dados dos usuários comprometidos para uso político dos dados em massa para influenciar as eleições americanas. Foram usados os resultados do teste e os dados do Facebook para construir um algoritmo que pudesse analisar perfis individuais do Facebook e determinar traços de personalidade ligados ao comportamento eleitoral (Cadwalladr; Graham-Harrison, 2018 apud Prado, 2022, p. 279).

Desde 2016, na sua gênese, a expressão *fake news* (FN) ganhou popularidade na boca do então presidente dos Estados Unidos, Donald Trump, quando este não aceitava críticas da imprensa e dava às críticas o nome de FN na intenção de abafar as denúncias de suas falcatruas recorrentes. Desde que o submundo desinformativo começou a se alastrar, pesou sobre os jornalistas a pecha de que eles eram os responsáveis pelas FN, impondo, portanto, uma semelhança, ou melhor, uma desdiferenciação, entre as FN e as notícias. Em meio à confusão, as pessoas comuns, no calor das emoções, não conseguem, muitas vezes, diferenciar a informação produzida por profissionais sérios daquela produzida sem procedimentos éticos jornalísticos.

“Antes de tudo, a expressão *fake news* não pode ser traduzida ao pé da letra, porque se são *fakes* não são *news*, pois, no jornalismo, em princípio, não existe notícia falsa, tanto que uma das normas é a checagem dos fatos antes da publicação, conforme evidenciado em estudos anteriores” (Prado, 2021, p. 48). “Mas é assim que são conhecidas as mensagens fraudulentas proliferadas na atual era da pós-verdade pela qual o mundo vem passando

2 Cambridge Analytica foi uma empresa anglo-estadunidense de consultoria política, criada por Robert Mercer em parceria com Steve Bannon, o estrategista-chefe da campanha de Donald Trump em 2016.

de forma descontrolada. Portanto, por motivos óbvios: para um fato se tornar notícia, a prioridade, entre as várias regras éticas da imprensa, é que ele seja verdadeiro, isto é, uma verdade factual – expressão cunhada por Hannah Arendt (1906-1975) no ensaio ‘Verdade e política’, publicado na revista *New Yorker*, em 1967” (Prado, 2022, p. 25).

Portanto, de modo geral, notícias e informações são confundidas com FN. Nesse contexto, não é pertinente chamá-las de “notícias falsas”, nem de “informações falsas”; é preferível chamá-las de “mensagens falsas” que, à luz de Capurro, podem bem ser denominadas “antiangélicas”. Além disso, foi ficando cada vez mais claro que a fábrica da desinformação abarca não somente as FN ou mentiras em estado puro – feitas com a intenção de enganar e prejudicar –, mas também uma distribuição de variedades que poderiam ser chamadas de pseudoangélicas, responsáveis pela generalização da desinformação, como se segue:

“1) informação descontextualizada, ou seja, sem a data, o que faz com que algo dito no passado (mesmo sendo verdade) volte à tona desatualizado, requentado; 2) mensagens adulteradas, manipuladas, editadas com recortes de modo a distorcer o contexto; 3) má informação, sem apuração ou mal apurada, ausência de fontes (muito menos de especialistas) para corroborar o que é dito, sem visualização transparente de dados fazendo com que qualquer opinionismo possa emplacar, mesmo com fontes forjadas; 4) mescla de verdade e mentira para deixar dúvidas e confundir; 5) sensacionalismo, aumentando e maquiando o teor do que é dito; 6) boatos, *hoax*, factoides, fatos alternativos etc., ou

mesmo de dados enviesados. São, no fundo, mensagens fraudulentas com algum grau de fragilidade e ‘com a carga que trazem, em geral, perspicaz (por conta da penetração de vista), porque geram todo tipo de emoção (boa ou ruim, dependendo da crença de cada um), fazem com que grande parte não desvie o olhar, tamanha é a atratividade” (Prado, 2022, p. 27).

Contudo, como um passo estridente a mais, a IA generativa atropela o rumo e o software GPT não é de hoje que traz alvoroço, mas desde final de 2022 causa sensação. Porém, por total falta de transparência das plataformas (incluindo as de mensageria³ conversacional) sobre o processo de circulação informativa, estamos todos nas mãos de quem manipula a ordem do conhecimento, da troca de saberes ou mesmo de *hard news* diárias que importa saber (Prado, 2022, p. 293). Pensemos em um dos exemplos de modelos linguísticos de IA generativa: o LaMDA, um gerador de diálogos, como um *chatbot*⁴, bem sofisticado, do nosso tempo. Antes, em 2019, o surgimento pela empresa

3 A principal ferramenta utilizada para a propagação de desinformação no Brasil é o WhatsApp. Em fevereiro de 2020, a pesquisa sobre mensageria móvel “Panorama Mobile Time/Opinion Box” revelou que o aplicativo está instalado em 99% dos smartphones brasileiros. Entre os usuários da ferramenta, 98% disseram que acessam todos os dias ou quase todos os dias (Galassi, 2021).

4 Um exemplo claro das novas funções das ferramentas nas redes sociais são os *bots*, que, quando usados com FN, espalham demasiadamente mensagens errôneas. Ao aprender – profundamente ou não – com a máquina, *bots*, *chatbots* e ciborgues incrementam esses desenvolvimentos invasores, ajudando a falsear, replicar e viralizar nas redes conteúdos de interesse exclusivo, produzidos com rigor minucioso de acordo com o resultado da análise dos dados, para direcionar de forma algorítmica determinados internautas, desde que sejam influenciáveis (Prado, 2019, p. 70).

OpenAI (apoiada pela Microsoft) do GTP-2, que gera de forma automática amostras de texto sintético, deixou muitos dos pesquisadores estupefatos, principalmente aqueles que lidam com o combate das mensagens fraudulentas. Afinal, tratava-se de ter à disposição um programa que produz *deepfake* de texto com uma velocidade impressionante (Prado, 2022, p. 326). Bom frisar que não há privacidade por design.

“O GTP-2 foi treinado em um conjunto de dados contendo cerca de dez milhões de artigos, selecionados no site de notícias sociais *Reddit* para links com mais de três votos. A coleção de textos pesava 40 GB. Na definição do *Blog OpenAI*, trata-se do treinamento de um ‘modelo de linguagem não supervisionada em larga escala que gera parágrafos coerentes de texto, obtém desempenho de ponta em muitos parâmetros de referência de modelagem de idioma e realiza compreensão rudimentar de leitura, tradução automática, resposta a perguntas. O objetivo é simples: prever a próxima palavra, considerando todas as palavras anteriores em algum texto’” (Prado, 2022).

Foi dito na ocasião que o modelo de IA, o GPT-2, era “tão bom e o risco de uso malicioso era tão alto” que tiveram de romper com sua prática habitual de liberar a pesquisa na íntegra para o público, explicou Herne, no *The Guardian* (2019). Ainda Vincent, no *The Verge* (2019), alertou que a motivação da OpenAI para restringir o lançamento incluía o potencial de programas como esse de criar “artigos de notícias enganosos” além de “automatizar spam e abuso”, como relatado no blog *Sociotramas* (Prado, 2019).

Já em 2020, quando o software ganhou nova versão, o GPT-3, treinado no texto de milhares de livros e na maior parte da internet, causou mais estardalhaço não somente por gerar textos “parecidos” com fontes escolhidas e no mesmo estilo, o que já assustava a todos, mas também pela possibilidade de gerar FN, na medida em que, desventuradamente, se pode seguir a linha de impostores que fabricam todo tipo de desinformação (Prado, 2019, p. 330). Agora, não dá mais para destreinar.

Como sabemos, trata-se da criação de softwares que integram a tecnologia de *chat*, no estilo robô, com os recursos do *big data*, com o poder de computação dos próprios dados digitalizados e disponíveis na infosfera. O desenvolvimento é a consequência da análise e moderação de conteúdo, seu armazenamento e da posterior raspagem para construir um molde forjando informações aglutinadas, no intuito de distribuição a partir de perguntas ou pedidos dos usuários. Claramente, se o pedido parte de uma mensagem falsa, o mecanismo vai identificar em páginas de busca e gerar a resposta falha promovendo mais fraudes, ou, no mínimo, dados falso-positivos. Pela falta de explicabilidade e opacidade algorítmicas, não sabemos quais dados incidem no aprendizado de máquina profundo, portanto, o risco de detonar a reputação é evidente!

Todas essas etapas desequilibradoras da informação correta – em prol de interesses, de modo geral, escusos, pois estamos falando de FN – passam a ser aceleradas em escala muito superior ao que víamos antes da possibilidade de uso do ciberespaço com as técnicas de propaganda para esse novo tipo de ações perturbatórias, e fazem surgir a desordem informacional com ou sem

chatbots. Nos tempos atuais, a propaganda, especialmente no campo político (com forte empuxo do marketing digital), sofisticou-se a ponto de tomar para si a responsabilidade de dar a base do alastramento das FN (Prado, 2022, p. 211).

Os esquemas geradores da pervasividade das FN de hoje, em pleno século XXI, beberam na fonte de funcionamento da propaganda tradicional. Entretanto, com os recursos tecnológicos, obviamente, elevaram todas as fases do processo de modulação do público selecionado. São eles: coleta e armazenamento de informações – mais ágeis e servindo-se de grande quantidade de dados (*big data*); análises destinadas a refinar, escolher o público a ser manipulado (aqui, neste recorte, pelas FN) – com acesso cada vez mais fácil proporcionado pelos próprios perfis, que caem em ciladas variadas: de concordância nos termos de condições de uso, passando pelo aceite de *cookies* de rastreamento, até por responderem testes de personalidade e comportamento que inundam as redes – e direcionamento, espalhamento e circulação – e que, com o auxílio de robôs, provocam disparos em massa, replicam e viralizam perfis falsos (Prado, 2022, p. 210).

Enquanto a sociedade – e, em especial, a comunidade jornalística, que recebe farpas de todos os lados, como se estivesse atuando no epicentro do problema da desinformação – estiver à mercê do agravamento dos sistemas de vigilância e moderação de informação baseados em algoritmos de IA para interferir e modular o pensamento de um público designado (preferencialmente vulnerável), estaremos de mãos atadas no sentido de ter um freio perante essa intrusão, que vem de sites especializados em FN (por vezes sensacionalistas), mas, especial-

mente, das redes e suas mais rentáveis plataformas – Facebook (com ele, WhatsApp e Instagram) e Google (inclusive YouTube) –, com milhões de acessos por segundo, nas quais grande parte das pessoas está e que passa por uma era de distribuição (de um lado, com controle de visualizações a seu bel-prazer e, de outro, totalmente desordenada de informações, porque qualquer absurdo advindo das brigadas da web tem lugar a ser publicado).

Em suma, um maior poder computacional reforça essa cola da crescente disponibilidade de grande quantidade de dados e o progresso dos algoritmos. Se pensarmos que, a partir de comandos de textos, a overdose de programas como o DALL-E 2, o Midjourney ou o Stable Diffusion gera imagens a partir de comandos verbais, ficamos à mercê desses resultados de IA, especialmente quando o uso é para o campo artístico.

Na internet dos humanos, muitos de nossos direitos e liberdades dependem das mil e uma questões do regime de atenção – olho por olho, tela por tela – do dilúvio das mensagens (com ou sem o gatilho do sentido) e do infopoder da computação, tal como o rastreamento de nossos dados em modo ininterrupto, efetivando de quando em quando a IA e com ela a fissura da mestria algorítmica.

CONSIDERAÇÕES FINAIS

Por mais que a discussão teórica aponte a necessidade de trocarmos as expressões *fake news*, “informações falsas” e “notícias falsas” por “mensagens falsas”, ainda demora para termos uma noção aproximada do flagelo infundido pela desinfor-

mação a nos entulhar cotidianamente, bem como das sequelas em seu rastro, embaralhando a noticiabilidade legítima e a informação que informa. Ao que parece, essa situação se estenderá *ad continuum*, pois toda e qualquer atitude para sanar esse fenômeno de anomalia informacional não é mais suficiente, tamanha é a desconfiança nas mensagens informativas. Hoje a internet amplifica o volume de desinformação, escalando o número de envolvidos; a IA traz alta velocidade de propagação e conseqüente viralização com *bots* e disparos em massa, facilitando a manipulação com *deepfakes*, cada vez mais bem-feitas tecnicamente e, portanto, ilusórias a olhos e ouvidos nus. No efeito de sucessão é a vez dos algoritmos, os quais, mediante a possibilidade da arquitetura da escolha de direcionamento por meio de monitoramento das manchas de informação dos usuários (e, por tabela, de seus pares), com o pente-fino da extração desses dados (emocionais, biométricos, georreferenciados etc.), mediante também a predição de padrões comportamentais, sequenciam raspagens na segmentação algorítmica com o intuito da personalização e formação das bolhas de público-alvo – aquele que se quer atingir – e, como os vestígios de dados – tão brutos quanto

posteriormente refinados da produção social – provocam outras camadas de dados, é possível fazer correlação para influenciar as próximas ações desses *clusters* para modular pensamentos e, conseqüentemente, comportamentos, exatamente de acordo com fins ideológicos e interesses (escusos) no que querem incutir.

O conceito de “sociedade da mensagem” e a adoção do termo “mensagem” ao nos referirmos às mal chamadas *fake news* podem, assim, iluminar a compreensão das dinâmicas de desinformação presentes atualmente no ambiente das redes sociais digitais. Ao menos, alivia a confusão que o público leigo faz ao determinar que toda e qualquer FN é informação ou notícia falsa. Claro que sabemos que em qualquer profissão existem desvios e, sim, alguns jornalistas erram e erram feio. Mas não é a regra.

Contudo, por sabermos que é a partir da filtragem em meio à arrebentação de conteúdo utilizada com técnicas de *pipelining* por desinformantes experientes nos estudos de modelagem da informação, especialmente em mensageiros instantâneos que, por serem criptografados, se espalham como um enxame que se move conforme o ambiente, tanto por isso será de difícil mitigação.

REFERÊNCIAS

- ABBAGNANO, N. *Dicionário de filosofia*. Trad. Ivone Castilho Benedetti. 6ª ed. São Paulo, WMF Martins Fontes, 2007.
- ANDERSON, C. W.; BELL, E.; SHIRKY, C. "Jornalismo pós-industrial: adaptação aos novos tempos". *Revista de Jornalismo ESPM*, n. 5. São Paulo, abr.-mai.-jun./2013, pp. 32-89.
- ARAÚJO, C. A. Á. "Correntes teóricas da ciência da informação". *Revista IBICT*, v. 38, n. 3. Brasília, set.-dez/2009, pp. 192-204. Disponível em: <http://revista.ibict.br/ciint/article/view/1240>.
- BARCELLOS, Z. R. et al. "Jornalismo das coisas". *Anais do 40º Congresso Brasileiro de Ciências da Comunicação*. Curitiba, Intercom, 2017. Disponível em: <https://bit.ly/38qWkV7>.
- BORKO, H. "Information science: what is it?". *American Documentation*, v. 3, n. 5, 1968.
- BUCCI, E. *Existe democracia sem verdade factual?*. São Paulo, Estação das Letras e Cores, 2019.
- CADWALLADR, C.; GRAHAM-HARRISON, E. "Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach". *The Guardian*. 17/mar./2018. Disponível em: <https://bit.ly/3DrpOgM>.
- CAPURRO, R. *What is angeletics?*. 2010. Disponível em: <http://www.capurro.de/angeletics.html>.
- CAPURRO, R.; HJÖRLAND, B. "O conceito de informação". *Perspectivas em Ciência da Informação*, v. 12, n. 1, 2007. Disponível em: <http://handle.net/20.500.11959/brapci/33134>.
- CASTELLS, M. *O poder da comunicação*. Trad. Vera Lúcia Mello Joscelyne. São Paulo/Rio de Janeiro, Paz e Terra, 2015.
- GALASSI, V. "Fake news: livro traz propostas para o enfrentamento real do problema". *CUT*. Distrito Federal, 9/jun./2020. Disponível em: <https://df.cut.org.br/noticias/fake-news-livro-traz-propostas-para-o-enfrentamento-real-do-problema-9ea7>.
- GIBSON, W. *Neuromancer*. São Paulo, Aleph, 2008.
- JUNGK, I. "Simondon: uma perspectiva ontoepistemológica para a contemporaneidade". *Transobjeto*, 2017. Disponível em: <https://transobjeto.wordpress.com/2017/04/15/simondon-uma-perspectiva>.
- LATOURETTE, B. *Jamais fomos modernos: ensaio de antropologia simétrica*. Trad. Carlos Irineu da Costa. São Paulo, Editora 34, 2005.
- MARKOWSKY, G. "Claude Shannon: american engineer". *Britannica*, 1/jun./2023. Disponível em: <https://www.britannica.com/biography/Claude-Shannon>.
- PRADO, M. "Algoritmos de geração de textos do GPT-2 da OpenAI tem seu código finalmente liberado – A comunidade agradece". *Sociotramas*. 9/dez./2019. Disponível em: <https://sociotramas.wordpress.com/2019/12/09/algoritmos-de-geracao-de-textos-do-gpt-2-da-openai-tem-seu-codigo-finalmente-liberado-a-comunidade-ai-agradece/>.
- PRADO, M. "Inteligência artificial na cultura informativa e algoritmos de enganação", in L. Santaella (org.). *Inteligência artificial & redes sociais*. São Paulo, Educ, 2019, pp. 57-72.
- PRADO, M. "La credibilidad periodística en jaque: conexión entre propaganda y fake news". *Ámbitos – Revista Internacional de Comunicación*, n. 53, 2021, pp. 216-30. Disponível em: https://institucional.us.es/revistas/Ambitos/53/Art_12.pdf.

- PRADO, M. *Fake news e inteligência artificial: o poder dos algoritmos na guerra da desinformação*. São Paulo, Almedina Brasil, 2022.
- RHEINGOLD, H. *Multitudes inteligentes: la próxima revolución social*. Barcelona, Gedisa, 2004.
- RHEINGOLD, H. *Smarts mobs – The next social revolution*. Disponível em: <http://www.smartmobs.com/>.
- SIQUEIRA, A. H. *Arquitetura da informação: uma proposta para fundamentação e caracterização da disciplina científica*. Tese de doutorado. Brasília, Universidade de Brasília, 2012. Disponível em: <https://repositorio.unb.br/handle/10482/12157>.
- SOUSA, J. P. *Elementos de teoria e pesquisa da comunicação e dos media*. 2ª ed. Porto, Porto Editora, 2006.
- WIENER, N. *Cybernetics or control and communication in the animal and the machine*. 2ª ed. Nova York, MIT Press, 1961.

Temidas, rebeldes, poderosas – Milênios antes do #metoo, elas davam as cartas

Marília Fiorillo

H

atshepsut, ou melhor, o *faraó* Hatshepsut (como ela preferia ser chamada) reinou como homem durante 20 anos na 18ª di-

Este híbrido de história e ficção é devedor de algumas viagens ao Egito, inicialmente para vasculhar o Museu Copta do Cairo e pistas dos manuscritos de Nag Hamadi, que contêm o mais antigo texto do cristianismo primitivo, o Evangelho de Tomé. Daí para o não planejado foram alguns quarteirões e viagens de vans ou felucas (o barquinho usado pelos camponeses para navegar no raso Nilo), visitas repetidas a museus, mesquitas e templos, que instigaram encantamento e uma posterior leitura e pesquisa bibliográfica. Sem esquecer do gentil intermediário de um visto episcopal brasileiro para que eu acessasse aqueles documentos sigilosos da Biblioteca do Vaticano vedados ao público.

MARÍLIA FIORILLO é professora de Filosofia Política e Retórica da Escola de Comunicações e Artes (ECA) da USP e autora de, entre outros, *O deus exilado – Breve história de uma heresia* (Civilização Brasileira).

nastia, idos de 1500 a.C. (antes da Era Comum). Foi um período de paz e prosperidade. Seus sucessores imediatos no comando do Egito se empenharam em um minucioso e tenaz trabalho de destruição de qualquer vestígio de seu reinado, nos relatos, estatuária e monumentos, esmerando-se em apagar seus rastros, numa operação de *piazza pulita* (terra arrasada) sem precedentes. Foi só no século XX, quando os norte-americanos escavaram em Luxor, que Hatshepsut reemergiu em sua formidável grandeza, vide o esplêndido templo em Luxor, aberto à visitação apesar de alguns ataques terroristas, e o restauro de inúmeros obeliscos e estátuas semidemolidos, raspados e vandalizados em Luxor e Karnak. A tentativa de aniquilá-la, removendo sua memória através da destruição de indícios materiais de sua existência, deu com os burros n'água. Ela tornou-se mais popular.

Algo parecido aconteceu com a filósofa, matemática e astrônoma neoplatônica Hipátia (c. 350/370-415 da Era Comum), na época em que Alexandria havia se transformado na “nova Atenas”. Foi assassinada com requintes de sadismo por monges cristãos (instigados pelo bispo ortodoxo, depois promovido a santo, Cirilo de Alexandria), que também destruíram a quase totalidade de sua obra. Os fanáticos sequazes de Cirilo, porém, exímios em retalhar a carne feminina, mas pouco espertos, se esqueceram de lançar ao fogo a farta correspondência de Hipátia com o bispo cristão-helenístico Sinésio de Cirene (atual Líbia), seu discípulo, amigo e admirador, além de intelectual versado em filosofia grega (Grazya, 1989). Muito do que se sabe dela provém

dessa amorosa troca de cartas – como as multidões que ela atraía para ouvir suas aulas, ou que era a principal conselheira do prefeito da cidade, Orígenes, ou ainda o quanto era indiscriminadamente querida tanto por pagãos como por neófitos cristãos não fundamentalistas. Além dessa fonte, há fragmentos de escritos de Hipátia em conjunto com o pai, o matemático Theon (que dirigia o Museu alexandrino), preservados em uma sala especial do Vaticano, e cujo acesso só é permitido com uma carta de recomendação e aquiescência episcopal. O assassinato de Hipátia foi um escândalo à época, a ponto de gerar uma tentativa de sindicância do mentor Cirilo (que deu em nada). Rendeu-lhe, porém, fama, difamação e honrarias póstumas nos séculos seguintes. Há inúmeros, vários livros sobre ela¹, a começar pelo *best-seller* de Charles Kingsley de 1853 (traduzido em sete línguas) e culminando, para um público maior, no monótono filme *Ágora*, de 2009, dirigido por Alejandro Amenábar e com Rachel Weisz interpretando a aparente doçura da filósofa². Hipátia tornou-se uma espécie de ícone do feminismo *avant la lettre*, o que provavelmente a desagradaria bastante, pois nem ela nem seus admiradores davam muita bola para questões de gênero. Talvez até ficasse desconfortável com esse epíteto “mulher & filósofa”. Ela se perguntaria se há uma lógica estritamente masculina (os incontornáveis silogismos) oposta a

1 Em especial os de Fawzia Assad (2000) e Joyce Tyldesley (1998).

2 Destaque para o livro de Margaret Alic (1989).

uma sofística tipicamente feminina (já que as mulheres são arteiras enganadoras...). Hipátia poderia mesmo se ofender se a reduzissem a tal clichê. Ela foi um pensador (substância) do sexo feminino (acidente). O que, pasmem, não causava estranheza nem inspirava militância há cerca de 20 séculos.

Já a sultana, ou melhor, o *sultão* feminino Shajarat al-Durr (o título “sultão Shajar” está inscrito em um dinar, moeda da época), foi outra governante egípcia que comandou exércitos no século XIII, durante a 7ª Cruzada, e derrotou os cristãos invasores. De origem armênia, provavelmente, foi vendida como escrava-menina para Al Salih Ayyub, com quem se casou mais tarde, quando ele se tornou sultão. Com a morte do marido, no auge do conflito e com o risco de colapso do Egito muçulmano, Shajarat tomou o lugar dele, escondida na tenda onde ocultava o cadáver, para que a notícia não se propagasse e desse ânimo aos inimigos, em 1250. Poucos sabiam que era ela que traçava as estratégias que encurralaram e aniquilaram os invasores. Em um ano, restituiu o Egito aos legítimos donos e despachou Luís IX de volta à França. Mas os emires aiúbidas e o califa sírio abássida não aceitaram curvar-se à nova sultana/sultão. Shajar então casou-se pela segunda vez com o novo governante do Egito, Aybak, mas continuou a comandar o país, dos bastidores. Anos mais tarde, ao perceber que o marido estava lhe escapando, mandou matá-lo durante o banho. Os mamelucos (outra facção do Islã de então) a protegeram, libertaram da prisão e evitaram sua con-

denação por assassinato. Mas ela acabou morta, em 28 de abril de 1257, a mando do filho adolescente de Aybak, da maneira mais extraordinária: a tamancadas, espancada pelas escravas do harém. Seu corpo nu foi arremessado fora dos muros da cidade. Seu mausoléu, uma pequena pérola arquitetônica, estava abandonado, infestado de mato e quase em ruínas nos anos 1990.

Há muito mais a contar, uma infinidade de peripécias históricas, sobre esses três indivíduos do sexo feminino que regeram, lideraram militarmente e educaram o Egito. Aqui, apenas lhes demos uma voz inventada. Que falem uma certa Hatshepsut, alguma Hipátia estraçalhada por desvairados monges cristãos e uma Shajarat, fustigada até a morte por outras mulheres. Mulheres, ressaltamos.

I. HATSHEPSUT

Eu, rei Hatshepsut, irmã e esposa de Thutmosis II, concebida por Amon, a mais amada dos filhos de Thutmosis I e Ahmose, de linhagem divina e sangue real, cujo nome e mando alcançaram até a distante Etiópia, cujo selo e governo trouxeram prosperidade e paz por 22 anos às margens do Nilo, cujos feitos, de tantos e tão magníficos, estão inscritos no mais alto obelisco de Karnak, para que nele coubessem as minúcias de minha opulência, eu, cujo templo mortuário foi erguido para empalidecer todos os palácios e templos e santuários do passado e do futuro, o solar e límpido Deir al-Bahri, flor arquitetônica incrustada no deserto, hierático, simétrico só a mim em esplendor e

nobreza, eu, Maatkare Khnemet-Amon Hatshepsut, soberano coroado, senhor do Alto e Baixo Egito, cujo nome ecoa como uma brisa seca, cujo selo ostenta o leão, cujos feitos estão adiante dos de qualquer geração, eu, rei e faraó, rei e regente, rei e consorte de mim mesma, eu, Hatshepsut-Amon, que envergo as vestes e a barba real, daqui de Tebas, no ano 21 da 18ª dinastia, escrevo a Senemute, meu amigo, arquiteto, amante e conselheiro, para exaltá-lo:

“O portal de sua casa estava aberto.
Meu bem-amado inclinado aos pés de sua
[mãe,
Irmãos e irmãs o circundavam.
E aqueles que passavam pelo caminho
Eram tomados de amor por ele,
Jovem perfeito e sem igual, de raras virtudes.
Ele pousou seu olhar em mim,
Pois eu o havia notado.
Quando penso no bem-amado
Meu coração se sobressalta
E confunde meus gestos.
Esqueço de vestir-me como deveria,
Negligencio meus leques,
Não maquio meus olhos,
Não me perfume mais com suaves aromas.
Oh, coração, não me exponha a tais penas.
Por que se conduz como um louco?
Venha para sua morada, bem-amado.
Você não tem inimigos.
Ó bela criança, venha para sua morada, para
[que possa me ver.
Sou sua esposa, aquela que o ama.
Não se afaste de mim, belo adolescente,
Venha agora para sua morada.
Meu coração pede por você, meus olhos o
[desejam.
Ah, que maravilha vê-lo, bem-amado.
Na cabeceira de minha cama

Possa você dormir, as narinas cheias de
[alegria,
E, amanhã cedo, despertar com Amon”³.

II. HIPÁTIA

Eu, Hipátia, filha de Theon, o guardião da Biblioteca alexandrina, filha da Ideia e irmã dos diversos saberes, instruída nas artes e ciências de Platão, Plotino e Ptolomeu, de linhagem grega no espírito e macedônica no sangue, eu, que interrogo o movimento dos céus, do Sol e das estrelas e por isso inventei o astrolábio, que pondero a gravidade de toda substância líquida e para isso inventei o hidrômetro, eu, astrônoma, matemática, geômetra, estudiosa do Cosmos e das emanções em que se degrada o Real, eu, presença que é como um magneto e atraí multidões para os salões da Biblioteca, sempre mais gente a me ver e ouvir, eu, cujas lições graves e serenas encantam a todos, judeus, romanos, gregos e egípcios do Delta, eu, cuja palavra esparge um fármaco que cura as exasperações, cuja fama corre o Mare Nostrum e fez de Orestes, o prefeito da cidade, meu ouvinte cativo, meu aluno mais próximo, eu, cujos conselhos têm o vigor da persuasão e força da autoridade, eu, que dobro os desígnios com o sopro do verbo e cujas exortações são comoventes e convincentes, eu, a última representante da filosofia helenística, vejo sobressaltada um mundo que está prestes a ruir e doravante abominará o atrevimento do intelecto ao curvar-se a um deus ciumento e

3 Colagem extraída de *Chants d'amour de l'Égypte Ancienne* (1996).

exclusivista. Eu, Hipátia, daqui desta nova Atenas, a gigantesca cidade de Alexandria, no ano 415, escrevo a Sinésio, fraterno e devotado aluno, que, soube, foi nomeado bispo da Cirenaica, de todo o norte da África, para aquietá-lo. Pois dele recebi esta carta de lamento e angústia:

“[De Ptolemaida a Alexandria, início de 413] Saudações, bem-aventurada Senhora, a ti e aos felicíssimos companheiros. Já há algum tempo pretendia repreendê-la por não me escrever, pois não me considera digno de uma resposta. E se tu, beata Senhora, e todos vós, me desdenhais, não será por minha culpa, pois não há culpa em ser desafortunado como só um homem pode ser. Mas se ao menos pudesse ler tuas cartas e saber como estás (espero que gozando da melhor fortuna), me bastaria, pois me regozijaria por ti, reduzindo assim à metade as minhas agruras. Porém agora o teu silêncio se junta aos males que me afligem. Perdi os filhos e os amigos, e a benevolência dos semelhantes. Mas a perda maior é a falta que sinto de teu diviníssimo espírito, o único bem que esperava me restasse para ajudar-me a superar os caprichos da sorte e os engodos do fado” (Grazya, 1989).

III. SHAJARAT AL-DURR

Sou eu, Shajarat al-Durr, quem comanda milhares de homens e centenas de batalhas, de dentro desta tenda onde jaz meu marido morto. Eu, nascida escrava e nômade, tornada esposa e serva de Sahli Ayyub, usurpo agora sua voz e seu pulso, e governo através do véu. Por 90 luas decido a cada momento o

que farão os generais inquietos, que esperam, do lado de fora da tenda, minhas ordens, que eles pensam ser as de meu marido morto, e por 90 luas nesta artimanha acumulo vitória sobre vitória contra os infiéis, glória sobre glória. E depois coroei a mim mesma sultão de todo o Egito e reinei sem disfarce por 80 outras luas e tantas batalhas, até que o califa de Bagdá e outros emires enviaram contra mim seus guerreiros e ódio. Escolhi não lutar e não fugir, mas casar-me com o mais valente dos meus algozes. Assim o fiz e tornei-me esposa de Aybak, e através dele, de sua docilidade, por tantas outras ininterruptas luas governei. Por trás do véu continuei a imperar através de meu segundo marido, de quem ocultava todo segredo político valioso, até que intrigas e sua covardia inata o contaminassem e ele planejasse me afastar. Adiantei-me, e mandei assassiná-lo a tempo. Se antes havia me servido do cadáver de um marido, depois me vali da ânsia carnal de outro. Mas esta esfriara com a decisão de tornar como favorita uma segunda esposa. Antes que ele me substituísse no leito e no palácio, forjei sua morte. Sempre fora exímia na arte de matar e seus subterfúgios. A fúria de meus adversários, porém, ganhou ímpeto e adeptos, embora não haja entorpecido alguns fiéis servos, que me libertaram do cárcere e da tortura e me conduziram a uma torre onde estaria protegida, não fosse a cólera vingativa do filho de Aybak, o menino de 15 anos que suplantou meus mais aguerridos adversários e conseguiu, enfim, acabar comigo. Não usou espada, adaga ou veneno: apenas me entregou às mulheres de seu harém.

A mim, Shajarat al-Durr, o último grande líder da dinastia aiúbida, o mais feroz, sagaz e destemido nas campanhas militares,

implacável na ação e indiferente à clemência, a mim, Shajarat al-Durr, a “Árvore de Pérolas”, gêmea da coragem e astúcia do curdo Saladino, a mim, que subjuguiei cristãos, persuadi muçulmanos e sujeitei tantos à minha vontade, sobre mim recaiu o ódio e rancor das mulheres.

Eu, o único soberano mulher que já houve no Islã, sussurro minhas últimas palavras à fiel escrava, daqui de al-Qahira, que os forasteiros chamam Cairo, minutos antes de ser conduzida ao festim das concubinas.

“Ontem foi gestado o delírio de hoje,
[deste dia
E a indiferença, o triunfo ou o desespero
[de amanhã.
Celebremos! Pois não sabemos de onde
[viemos ou por quê.
Celebremos! Pois não sabemos por que
[iremos, nem para onde.
Quê! Um desatinado Nada não pode
[provocar o jugo.
É tão tolo ressentir-se de prazeres fruídos
[na proibição como temer castigos eternos
[diante da dor que despedaça o presente.”

IV. OS MUITOS MODOS DE MORRER

Em 1458 a.C, portanto, há 3.500 anos, assim que Thutmosis III, seu sobrinho, foi coroado, Hatshepsut foi submetida a uma segunda morte. Por ordem do novo faraó, de quem ela havia sido regente, foram destruídos todos os indícios de sua existência, monumentos e inscrições que lembrassem a antecessora. Alguns foram demolidos, raspados e reduzidos a fragmentos de granito ou calcário; outros, adulterados para que, no lugar onde antes aparecia a imagem de-

la, fosse entalhada a de Thutmosis III. A movimentação deve ter sido intensa, pois os arquitetos de Hatshepsut haviam erguido inúmeros monumentos, naturalmente ornamentados com a efígie da soberana-faraó, e sua supressão só pode ter exigido um empenho de destruição comparável ao da criação. Não bastasse ter sumido da pedra, o faraó Hatshepsut também desapareceu dos papiros e foi removido das listas dos cronistas da história do Egito (apenas um deles, Manetho, citado pelo historiador judeu Flavio Josefo, registrou sua passagem). Para os escribas oficiais, porém, sempre constou que o reinado de Thutmosis I foi seguido imediatamente pelo de Thutmosis III. Hatshepsut Maatkare, aquela que se intitulou rei, permaneceu um fantasma até o começo do século XX, quando arqueólogos da equipe de Herbert Winlock, do Metropolitan Museum de Nova York, casualmente desenterraram, na vizinhança do Deir al-Bahri, uma considerável quantidade de fragmentos de imagens da rainha-rei, depois restaurados e hoje expostos no Metropolitan e nos Museus do Cairo e de Luxor. Não fosse esse feliz acaso, Hatshepsut, o mais importante governante da 18ª dinastia, continuaria ignorada pela posteridade. A morte simbólica de um faraó é bem mais grave que a extinção física – que é só uma passagem para uma nova vida, daí as tumbas serem recheadas de artefatos, móveis e joias a serem recuperados no despertar. O essencial era a sobrevivência na *outra margem* ou no reino dos mortos, mas para que isso se consumasse era mandatório que houvesse representações, nesta vida, da figura do morto. Thutmosis III, embora não tenha eliminado fisicamente sua antecessora,

tratou de se assegurar de sua verdadeira morte, isto é, de que ela jamais alcançaria a *outra margem* do rio, uma vez apagada no registro da história e memória. Ao extinguir seu passado, aboliria seu destino, garantindo que ela jamais passasse pela balança de Anúbis, que compara o peso do coração do morto ao de uma pena e decide seu futuro.

A morte de Hipátia foi um episódio de insânia excepcional até mesmo para os padrões esquisitos de culto ao martírio, praticado pelos eremitas e ascetas do deserto (hesicastas) que buscavam a salvação jejuando e mudos em cavernas por anos, conforme a “Filocalia”⁴. Hipátia foi retalhada até a morte por cacos de conchas, por uma turba de monges cristãos da Tebaida. Sua cabeça foi lançada ao mar e os nacos de carne, pele e ossos, queimados. Era um período, o mesmo de Agostinho de Hipona (depois santo, apesar de seus pendores cripto-protestantes), de conturbada transição entre o cosmopolitismo pagão romano, mais tolerante, e o intransigente e ressentido cristianismo eclesiástico. O ardiloso bispo Cirilo, em sua ambição de engolfar o poder, temia a influência da filósofa sobre os ricos e poderosos da cidade, sobretudo sobre Orestes, o prefeito. A versão da carnificina de Hipátia adotada por Gibbon (2010) é a de que ela foi retirada à força de sua liteira quando estava a caminho de uma de suas palestras públicas, despida, arrastada para a igreja

local e ali estripada pelos monges, uma agonia que deve ter sido longa, pois eles usaram pequenas conchas para separar a carne dos ossos. Em seguida, os monges a esquartejaram (alguns dizem que jogaram sua cabeça ao mar), queimaram o que havia restado do corpo e foram até Cirilo saudá-lo como libertador.

Quanto a Shajarat al-Durr, as esposas e concubinas do harém não desperdiçaram a oportunidade de vingar-se da favorita do sultão e a surraram até a morte, golpeando-a com seus tamancos. Uns dizem que o cadáver da única mulher-sultão do Islã foi pendurado no centro do Cairo para servir de repasto aos cães e diversão ao povo. Outros, que foi arremessado seminu, com um tecido de seda e pérolas amarrado à cintura, logo saqueado pelos passantes. Outra lenda, ou fato, é que alguém da equipe do arqueólogo Howard Carter comprou em 1903 uma caixinha de madrepérola no mercado Khan el Khalili e esqueceu-se dela. Anos mais tarde, seu neto abriu-a e encontrou um fino rolo de papiro que descrevia, no árabe mais requintado, as façanhas desta extravagante e potente mulher.

A mim, Hatshepsut, sequer um duplo foi concedido para cruzar a *outra margem*. Milênios de silêncio, até ressurgir no deserto. De mim, Hipátia, algo subsiste enredado nos corais do fundo do mar. De mim, sultão Shajarat, cujo epílogo foi servir de repasto aos cães, a história preservou uma efígie numa moeda.

Eu-Hipátia pergunto se teriam sido as mesmas mãos que esquartejaram meu corpo e despedaçaram meus livros, Eu-Shajarat suspeito que meus braceletes foram para a mais jovem concubina, Eu-Hatshepsut, por

4 A Biblioteca do Mosteiro de São Pacômio, no Egito, tem o maior e melhor acervo sobre o cristianismo ortodoxo dos cinco primeiros séculos, verdadeiras preciosidades. Ver: voskrese.info.

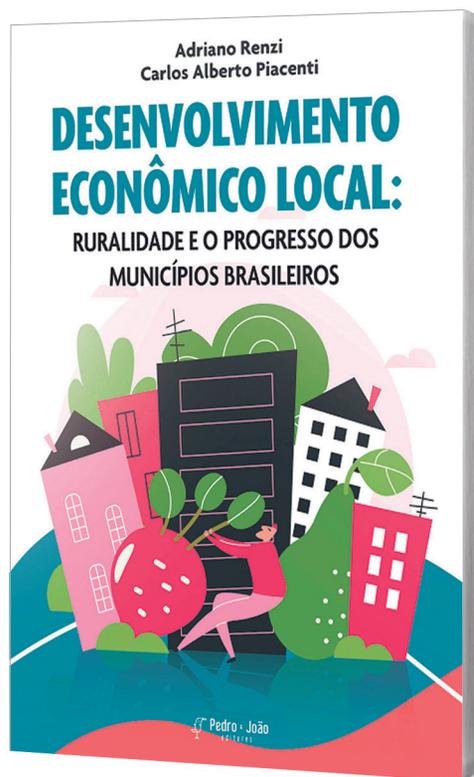
pouco escapei de tornar-me incógnito fragmento de pedra, Eu-Shajarat, que dobrei o Islã e a cristandade a meus desejos, Eu-Hipátia, que regia uma cidade infiltrando-me no intelecto de seus homens, Eu-Hipátia, a complacente e sábia, Eu-Hatshepsut, a

magnânima, Eu-Shajarat, inclemente, astuciosa e diante de quem todos tremiam de medo, Eu-Nós, amadas e temidas em vida, depois atiradas ao esquecimento, vandalizadas, párias, silhuetas, névoa, fulgurância negra: hoje, emergimos das sombras.

REFERÊNCIAS

- ALIC, M. *L'eredità di Ipazia: donne nella storia delle scienze dall'antichità all'Ottocento*. Roma, Riuniti, 1989.
- ASSAAD, F. *Hatshepsout, femme pharaon: biographie mythique*. Prefácio de Michel Butor. Paris, Librairie Orientaliste Paul Geuthner, 2000.
- CHANTS *d'amour de l'Egypte Ancienne*. Paris, La Table Ronde, 1996.
- GIBBON, E. *The history of the decline and fall of the Roman Empire*. Nova York, Gale Ecco, Print Editions, 2010.
- GRAZYA, A. (ed.). *Opere di Sinesio di Cirene*. Torino, Utet, 1989.
- TYLDESLEY, J. *Hatshepsut the female pharaoh*. Londres, Penguin Books, 1998.

livros



Semeando futuros e retocando passados

Daniel Afonso da Silva

Desenvolvimento econômico local: ruralidade e o progresso dos municípios brasileiros, de Adriano Renzi e Carlos Alberto Piacenti, São Carlos, Pedro & João Editores, 2023, 322 p.

Jorge Luis Borges alimentava a convicção que sugere que um bom livro é aquele que lê o seu leitor. Quem navega pelos prados das *belles-lettres* de alta expressividade e sofisticação nota fácil que o argentino tinha nisso muita razão. Italo Calvino, outro mestre desse tipo de concepção, do outro lado do Atlântico e desde alguma de suas *idades invisíveis*, ponderava na mesma direção; e ainda ia além, afirmando ser difícil sair de um livro formidável sem a sensação de que ele – o livro – e seu autor, ao fim das contas, parecem sorrir. Sorrir um sorriso de efeito caborteiro. Cheio de malícia. Coisa de Medusa. Cheio de charme. Coisa de Mefistófeles. Quase machadiano e cheio de miradas de soslaio. Tipo oblíquas, feito ressaca. E tipo circulares, feito provocação, do início ao fim. Ora no entusiasmo, ora em comoção. Ora em maravilhamento, ora em estupefação.

Em altas e belas letras – aquilo que Ovídio, Quintiliano, Horácio, Cícero, Vir-

gílio denominavam *littera* –, em maior ou menor grau, é – ou pode ser – sempre assim. Uma tempestade em primor. Um cataclismo ao resplendor. Um caminho sem volta ao conhecimento e ao saber.

Borges que o diga. Calvino que o demonstre.

Em outros gêneros, geralmente distantes da alta e altíssima sofisticação das *belles-lettres*, nem sempre essa métrica de ilustração se transforma em convenção. Em verdade, para ser justo e sincero, longe das áreas imperativamente regidas pelo peso moral da modéstia das Artes – com “a” maiúsculo e imaginação infinita – essa realização simbólica ocorre *quase nunca*. Nos escritos acadêmicos, por exemplo, nos últimos tempos, que beiram século, esse *quase nunca* deu lugar ao *quase sempre* para ser enfática e taxati-

DANIEL AFONSO DA SILVA é doutor em História Social pela FFLCH-USP e integrante do Núcleo de Pesquisa em Relações Internacionais (Nupri) da USP.

vamente *nunca*. Nunca ou jamais. Jamais ou nunca. Como se queira desejar.

Isso porque, desde o início da agudização da especialização da razão nos séculos primevos da Modernidade, os grandes filósofos, artistas e autores foram virando singelos acadêmicos. *Auctoritas* e autorais. Avessos à *imitatio* e ao *aemulatio*. Debochadores da *mimese* e do controle do imaginário. Blasfemadores do decoro e do cânone. E, nessa condição, foram transformando o saber e o conhecimento n'algo restrito, vazio, obscuro, sem charme. Cada vez mais útil – é bem verdade. Mas, cada vez, também, menos luminoso. E, por isso, desidratado sem expressão. Sem forma. Mesmo que eivado de conteúdo.

Nietzsche e Wittgenstein – para ficar apenas neles, maiores em suas épocas, e não rememorar afoita e desesperadamente Vico, Kant, Hegel, Schopenhauer, Weber, Keynes e mesmo Marx – passaram a vida se deblaterando contra essa conformação. Nenhum deles acreditava nem suportava a separação da ciência *versus* as artes, do saber elaborado frente à vida, do conhecimento estruturado diante da fé, dos rigores da universidade *versus* os imperativos convencionais de cada tempo. Todos eles defendiam um saber com obras e uma universidade com ação. Uma amputação desses fragmentos do Todo era, para eles e muitos outros, uma ignomínia. Uma indecência. Uma irresponsabilidade. Coisa de burocratas. Coisa de onanistas mentais. Quem sabe coisa de imaturos arlequins, histriões, truões, farsantes, bonifrates, *clowns*. Aqueles mesmos que Max Weber (1864-1920) apartou do *savant* – o intelectual, o professor, o sábio – e dei-

xou bem perto do *politique* – o sujeito avesso aos rigores da cientificidade da veracidade da razão.

Mesmo assim, a separação prosperou e avançou muito depois da derrubada massiva dos mantos purpúreos e das negras sotainas. Ou seja, depois das revoluções, de Napoleão e da burocratização, geralmente universitária, da produção de conhecimento e saber.

De toda sorte, o ocaso da Modernidade tudo isso revisitou.

Não foi nas trincheiras que esse ocaso começou. Mas foi nas suas ruínas que ele se acentuou. Notadamente no triênio 1942-1944 e no biênio 1944-1945. Nesses momentos, a Modernidade chegou ao seu estágio máximo de paralelismo, contradição, entropia e agonia. Auschwitz-Birkenau, campos de concentração. Eis o produto da Modernidade e de sua racionalidade. Se nada disso bastasse, vide o sadismo sem paralelos plasmado nas quedas de Berlim, Tóquio, Hiroshima e Nagasaki. Tudo, a rigor, fruto da Modernidade.

Sob a sua hoste, nunca, portanto, o ser humano fora tão conscientemente marginal. Vide as centenas de milhares de pessoas queimadas vivas nas ofensivas ocidentais de liberação. Lembre-se dos milhões de outros – 60, 70, 80 milhões – que tiveram a sua vida injuriada para nunca mais reabilitar.

Como seguir endossando essa Modernidade? Uma Modernidade de razão iluminista. Que especializava o saber, mas que retirava dele toda fé e todo o coração? Como seguir endossando tudo isso depois das monstruosidades das guerras totais?

Moralmente, não dá para endossar.

Adeus, então, Modernidade?

Se fosse assim, seria fácil.

Mas, não: não foi assim.

O arranjo encontrado foi a Pós-Modernidade. Com o prefixo “pós” bem situado para indicar o ancoramento do após no antes. Fazendo-se, assim, um eterno mal-estar. Da agonia da Modernidade ao mal-estar da Pós-Modernidade. Tudo porque o traumatismo da longa noite escura – leia-se: das guerras totais de 1914 a 1945 –, ao produzir o mal-estar da civilização, também vandalizou todas as referências modernas e, no contrapelo, foi incapaz de colocar outras no lugar.

Ocorreu, desse modo, nesses tempos de tormentas – e quase ninguém ousa mencionar –, um verdadeiro sinistro mental e moral eternamente irremediável. Nunca se viu tanta selvageria entre os homens nem tanta brutalidade entre iguais. E não apenas nos campos de batalha. Mas na vida em geral e como um todo. Desse modo, todos os campos da existência humana foram, assim, violados, violentados, deturpados. Como decorrência, nenhum corredor da produção de ciência, arte, conhecimento, informação e saber sobreviveu incólume. O impacto de tudo isso, portanto, sobre a universidade foi – e continua sendo – imenso. Todos os seus códigos e decoros de produção de conhecimento e saber foram, por tudo isso, gravemente lesionados.

Por essa razão – e não é de se estranhar – que, após o dilúvio de tragédias do apocalipse de 1914-1945, os anos que se seguiram deram vazão à afirmação do fim das certezas, das verdades, dos fundamentos universais, dos critérios absolutos. Era o último suspiro da Modernidade e a inauguração da Pós-Modernidade. Os

anos eram 1950-1970. Um verdadeiro *momentum*. O *momentum* de uma inclemente hecatombe epistemológica sem concessão. Onde nenhum paradigma suportou sem, ao menos, trincar e nenhuma área do saber permaneceu viva sem rachar. Era a crise geral de paradigmas. Onde nas ciências ficou em pé e, por isso, tudo precisou remoçar.

A gravidade disso tudo foi tamanha que inclusive a suspeição de Einstein sobre a disposição de Deus em jogar dados com o universo perdera a totalidade de sua relevância. Descobriu-se – ou se lembrou – amargamente e se popularizou que Deus sempre foi onisciente, indiviso e insondável. Sendo assim, por lógica, era impossível prever as suas jogadas tanto quanto a sua intenção em jogar. Dito francamente e sem parábolas, esse “simples” (re)descobrimto levou a pique todos os profundos fundamentos da física clássica que, a rigor, fundamentaram a ciência moderna, legitimaram a Modernidade e justificaram a especialização universitária do conhecimento e do saber. O rei, assim, ficou nu e as suas concubinas, incapazes de cobri-lo. A Teoria da Relatividade e a física quântica esmagaram de 80 a 90% de tudo que cientificamente se fez, imaginou e pensou ao menos desde Nicolau Copérnico (1473-1543) e Giordano Bruno (1548-1600) até chegar aos experimentos de Niels Bohr (1885-1962). Quem tiver alguma dúvida – por mínima que seja –, que volte com perícia aos feitos de Werner Heisenberg (1901-1976) e/ou de Robert Oppenheimer (1904-1967).

Dito sem retoque nem interjeição, a inquestionável hecatombe epistemológica da viragem da Modernidade para a Pós-

-Modernidade tornou Newton, Galileu e Descartes – distintos pilares da ciência moderna como um todo e da física clássica em particular – anacrônicos, estereis e sem razão. Assim como Darwin – e muito do materialismo de Marx – também.

Não dá para tomar consciência de tudo isso sem se espantar. Praticamente todo o saber humano fabricado sob a Modernidade precisou ser revisto e se revisar. Nessa, portanto, imoderada revisão, virou, assim, banal se reconhecer que o observador altera a natureza do objeto observado. Que o experimento da dupla fenda deu e dá a tudo isso razão. Que o bóson de Higgs, do recentemente falecido professor Peter Higgs (1929-2024), segue cientificamente uma verdade. Que infinitos corpos, assim, podem ocupar o mesmo lugar no espaço. Que existe o vácuo quântico. Onde tudo é onda e vai emaranhando. E que isso tudo é ciência e não cabala nem misticismo. Ciência nova. Nova ciência, sob novos paradigmas.

Sim: impressiona, emociona e desconcerta.

E desconcerta ainda mais porque o impacto de tudo isso sobre a universidade foi devastador. O *novo saber* sobre o conhecimento e o *novo conhecimento* sobre o saber simplesmente esquarteram todos os preceitos iluministas da cisão de ciência *versus* as artes, artes *versus* vida, rigor acadêmico *versus* fluidez de conhecimento e saber. Consequentemente, todos os seguimentos universitários precisaram internalizar essa nova realidade, esses novos tempos, essa nova forma de fazer e pensar ciência. O mal-estar gerado por tudo isso foi – e continua sendo – extraordinário. A incerteza tomou conta

de praticamente todas as áreas do saber de modo a ninguém nem nenhuma área universitária conseguir grassar em imobilismos. Era o *momentum* 1950-1970. Um momento rude e implacável para os produtores de conhecimento e saber. Um momento em que muitos simplesmente caíram em desespero por literalmente não conseguir saber o que fazer.

Que fazer?

No espaço universitário das ciências humanas e das humanidades – áreas, desde o Setecentos, tidas como “moles” por *manque* de aderência contundente aos preceitos de Verdade, com “v” maiúsculo, impostos pela ciência moderna – uma tentação de cientificidade tomou conta de praticamente tudo e de todas as áreas, desde a filosofia às letras à história à geografia à economia à sociologia à antropologia e afins. Nada, desse modo, escapou dessa tentação. Nem mesmo as convenções muito antigas feitas aquelas que, ao menos desde Heródoto, Quintiliano e Aristóteles, sempre entenderam que a filosofia, as letras, a história (entre outras) eram irmãs das artes, namoravam a totalidade e granjeavam abarcar fragmentos do infinito. Nada, mesmo dessas convenções, restou. O que causou avarias irremediáveis no interior das ciências humanas e das humanidades.

Discorrer sobre as particularidades dessas avarias importaria discussões exaustivas, em outros níveis e em outros lugares. Aqui e por agora talvez seja somente relevante acentuar que, do martírio da Modernidade ao frescor da Pós-Modernidade, a integralidade dos fundamentos ontológicos, epistemológicos e metodológicos de todas essas áreas foi violentamente ca-

nibalizada e, frente a isso, todas essas áreas, por razões existenciais, precisaram vivamente reagir. Dessa maneira e de modo sistematizado, o conjunto das ciências humanas e das humanidades se organizou em torno de duas estratégias: a) fugir da discussão ou b) garimpar mecanismos racionais de compensação.

Quem não fugiu, buscou refúgio num conceito chamado de interdisciplinaridade.

Ou seja, partindo-se dos imperativos da física quântica que aduzem que tudo é onda, tudo está emaranhado e múltiplos corpos podem ocupar o mesmo lugar no espaço, então tudo é interdisciplinar. E sendo interdisciplinar também poderia ser multidisciplinar e transdisciplinar.

Inter, multi e trans. Eis a estratégia de compensação e o lugar de refúgio dos praticantes das ciências humanas e das humanidades a partir do *momentum* 1950-1970.

Tudo bem. Muito interessante. Sem problemas.

Até que um baiano, academicamente experimentado, nascido em Brotas, formado em Salvador e tornado doutor em Strasbourg, na França, após viajar o mundo e contrastar consciente ou inconscientemente essa estratégia com variadas audiências, chegou a uma ponderação singular e desconcertante que afere o seguinte: não existe inter, multi e transdisciplinaridade sem disciplinaridade.

Ou seja, não existe relação entre disciplinas sem a fundamentação das disciplinas. Portanto, não existe diálogo disciplinar sem a afirmação da identidade disciplinar. Filosofia, letras, história, geografia, economia, sociologia, antropologia (e outras) dentro da universidade são áreas/disciplinas diferentes. Com códigos e

decoros diferentes. Que para seriamente se relacionar precisam, antes de tudo, se afirmar. Uma afirmação que só se realiza através *aggiornamento* de natureza ontológica, epistemológica e metodológica.

Dito sem meias-palavras, esse brasileiro, baiano, nascido em Brotas, dinamitou a estratégia de compensação empregada pelas ciências humanas e pelas humanidades ao evidenciar que o recurso ao refúgio da inter, multi e transdisciplinaridade era, em verdade, uma tentativa de fuga da afirmação ontológica, epistemológica e metodológica de cada uma das áreas.

O nome desse brotense ilustre é Milton de Almeida Santos (1926-2001) e essas suas ponderações simplesmente reabilitaram o mal-estar de se produzir saber e conhecimento depois do *momentum* 1950-1970. Sendo frio e direto, elas indicavam que ser academicamente honesto envolvia essencialmente uma meditação sobre método.

Quem seguiu relutando a essa meditação encontrou na noção de complexidade de Edgar Morin – nascido em 1921 e ainda vivo – outra estratégia de compensação. Ou seja, novamente partindo-se dos imperativos da física quântica que aduzem que tudo é onda, tudo está emaranhado e múltiplos corpos podem ocupar o mesmo lugar no espaço, então tudo é complexo.

Mas, novamente, não frutificou.

E, agora, como refúgio, os praticantes das ciências humanas e humanidades preferiram migrar para ilhas. Ilhas cada vez mais distantes e difíceis de se penetrar.

Nessas ilhas, as questões de método viraram mistérios tão bem guardados que nenhum estrangeiro – leia-se: alguém vindo de outra área – consegue decifrar. Desse modo, o diálogo sincero

e verdadeiro entre antropólogos, museólogos, economistas, historiadores, politólogos e afins virou quase impossível. O tipo de saber produzido por cada um deles ficou mais e mais sofisticado, complexo e acadêmico. Acadêmico, talvez, até demais. Tanto que os seus mecanismos de verificabilidade viraram praticamente insondáveis para quem não é da área.

Deixando essas tormentas de lado, o único aspecto aferível da produção de conhecimento e saber no interior das ciências humanas e humanidades que restou foi aquele da formalização dos resultados. Ou seja, da comunicação – leia-se: da escrita –, do conhecimento e do saber. E, nesse sentido, a Modernidade e a Pós-Modernidade contribuíram igualmente para o afastamento da ciência das artes. Ou seja, da comunicação por *belles-lettres*.

No mundo inteiro foi assim e no Brasil não foi diferente.

No plano geral, desde que dilemas gerais da viragem da Modernidade à Pós-Modernidade chegaram às universidades brasileiras que o martírio da cientificidade passou por aqui também a imperar. Praticamente toda a agonia de fora ganhou, por aqui, feições nacionais. Tome-se, por exemplo, o caso da economia – uma das áreas mais cientificamente relevantes do espaço das ciências humanas e das humanidades – e examine-se a sua relação com a comunicação de resultados.

Francamente, no horizonte de uma ou duas gerações após o *momentum* 1950-1970 chegar por aqui, pelos *campi*, o esforço intelectual da comunicação com precisão, alguma imaginação e parcelas de ironia foram esmaecendo até quase desaparecer. Nesse sentido, aquele sorriso matreiro das

frases de um Celso Furtado (1920-2004), por exemplo, começou a rarear e aquela ironia desconcertante das manifestações de um Roberto Campos (1917-2001) desapareceu para nunca mais voltar. Isso porque, adicionada à hecatombe epistemológica global, a inclemente pasmaceira de autoenganos que o *politically correct* universitário recentemente mobilizou acabou por ostracizar toda a relevância artística de toda essa gente. No mesmo diapasão, as convicções implacáveis – corretas e equivocadas – de uma Maria da Conceição Tavares viraram *démodés*. Isso porque, além da pasmaceira do *politically correct*, a tentação *woke* começou a literalmente lobotomizar trabalhos, autores e intenções. Por conseguinte, passou-se a se menosprezar o peso da retórica – e, portanto, das artes – no convencimento. Por tudo isso, aquele charme tipo *british* de um mestre feito John Maynard Keynes (1883-1946) passou a ser, no Brasil, quase condenado. De modo que a sensibilidade genuinamente literária de gente da qualidade artística de André Lara Resende, Pérsio Arida, Paulo Nogueira Batista Jr. e Eduardo Giannetti da Fonseca passou a ser tratada como “não ciência”, “empulhação”, “confusão”.

Impressiona, mas tem sido assim. E sendo assim, de uns tempos para cá, teses acadêmicas de mestrado ou doutoramento – e não apenas na economia – apresentadas nas universidades brasileiras começaram a aparecer sem tese. Se isso não bastasse, essas teses sem tese viraram pastagem fértil para plagiadores e indigentes universitários.

Dito e vendo dessa maneira – copo integralmente vazio –, dá-se a impressão de terra arrasada. É quase. Mas não

ainda. Ainda não porque, por mais que a interiorização das derivas epistemológicas mundiais sobre a realidade universitária brasileira tenda a dificultar novas ideias, novos atores e novas posturas, novidades ainda existem. Das muitas novidades materializadas em autores e obras disponíveis recentemente no Brasil, merece muita atenção, como exemplo, o doutor Adriano Renzi e o seu formidável *Desenvolvimento econômico local: ruralidade e o progresso dos municípios brasileiros*, supervisionado por Carlos Alberto Piacenti.

Pelo título – *Desenvolvimento econômico local* –, o livro indica, de saída, não ter nada que ver com *belles-lettres*, tampouco com assuntos amenos. Com o subtítulo – *ruralidade e o progresso dos municípios brasileiros* –, adicionam-se a essa fachada carrancuda ainda mais complicações.

Visto por esse aspecto, sim: o livro não é simples nem brando e aparenta habitar uma ilha. Mas, não. Basta abrir o livro e seguir para ver que não. O livro, malgrado técnico, não é enfadonho nem totalmente despido de ambições, emoções e arte.

Lê-se, assim, com gosto. Ou melhor: é-se lido por ele.

Do início ao fim ele apresenta um conjunto de provocações guiadas feitas conscientemente para perturbar o leitor. Impossível, desse modo, singrar atentamente de capa a capa sem perder o prumo.

Aprende-se muito lendo esse livro pois surpresas variadas vão emergindo página a página, seção a seção, capítulo a capítulo. Mas a surpresa decisiva começa a aparecer nos momentos finais da leitura, no qual se encontram os resultados originais da pesquisa. E é justamente nesses resultados que o livro e o seu autor parecem, enfim, sor-

rir. Restando ao leitor ser por eles lido. É incrível a sensação: *Desenvolvimento econômico local* e Adriano Renzi leem o leitor, sorriem por último e desconcertam quem acredita que domina a arte de sorrir.

Não se sabe se autor e obra fazem isso conscientemente ou não. Mas, ao fazê-lo, o livro e o seu autor induzem o leitor a um profundo exame de consciência, convencendo a todos da relevância da análise de dados, dos métodos quantitativos e da produção de índices e indicadores na contemplação segura da realidade brasileira. Tudo muito técnico – é verdade – e tudo muito conceitual. Mas, ao mesmo tempo, tudo muito bem pensado, medido e meditado nos planos ontológico, epistemológico e metodológico. Ninguém tem dúvidas tratar-se de um trabalho de economia. Ninguém tem dúvidas tratar-se de um assunto essencial. E ninguém tem dúvidas que existe nessa tese uma tese.

Trata-se, assim, de um esforço intelectual raríssimo entre os universitários brasileiros que envolveu fazer um estudo técnico, sério e duro, produzindo-se assim um livro audacioso, mas, ao mesmo tempo, um livro cheio de charme.

Saído de uma tese de doutoramento defendida em 2021 no Programa de Pós-Graduação em Desenvolvimento Regional e Agropecuário da Universidade Estadual do Oeste do Paraná, agraciada com o Prêmio Capes na categoria Planejamento Urbano e Regional/Demografia, retrabalhada num pós-doutoramento na Universidade Estadual de Maringá e tornada livro entre o inverno e o verão de 2023, na longínqua Universidade Federal da Grande Dourados, onde o doutor Adriano Renzi leciona Macroeconomia, esse livro tem,

portanto, uma longa e significativa história. Uma história que recobra outras histórias, como aquelas que envolveram os estudos também muito ousados e determinados que resultaram em “Desenvolvimento econômico: do crescimento à prosperidade” (Renzi; Henz; Rippel, 2019) e “Crescimento econômico nos municípios do Paraná-Brasil: uma análise com econometria espacial” (Renzi et al., 2022).

Por toda essa trajetória, o livro é eivado de rigores e convicções. Aqui, por certo, não é o melhor lugar para se explicitar as dimensões *positivas* e *normativas* desses rigores nem dessas convicções. Mas vale notar que existem e são importantes.

Indo ao coração do livro, ele foca os enigmas do desenvolvimento econômico em geral e do desenvolvimento econômico brasileiro em particular e se espraia por esforços de melhor mensuração da noção de *local* e maior calibragem técnica dos impactos do *novo rural* – a saber, da ruralidade – na compreensão da multidimensionalidade dos municípios brasileiros. Um assunto, de saída, difícil, complexo e engajado. E, mais que isso, um assunto presente em todos os debates verdadeiramente sérios sobre a realidade brasileira atual, pretérita e futura.

Nesse sentido, é de valor se ressaltar que a grande contribuição do livro reside na construção de um painel de indicadores com 146 variáveis de alta sensibilidade para a aferição de dimensões ambientais, demográficas, econômicas e sociais de 5.560 municípios brasileiros.

Isto: simplesmente isto.

Não precisa ser universitário tampouco economista para notar o rigor do enquadramento e a eficácia do método envoltos nisso.

Quem tiver pressa para tudo isso perceber, basta correr ao terceiro capítulo e vasculhar as páginas da 114 à 146. Lá existe uma tabela com todas as variáveis projetadas e todas as descrições dos seus fatores de aferição. Todas essas informações, em seguida – no próprio capítulo terceiro, mas principalmente nos capítulos seguintes –, vão espelhadas em dados do IBGE, notadamente do Censo e do Censo Agropecuário, do DataSus, do Ipeadata e do Atlas de Desenvolvimento disponíveis para os períodos de 2004-2006 a 2012-2013.

Feito, organizado e publicado, o imenso trabalho contido nessa tabela parece bem simples, trivial, truísmo, banal. Mas quem se detém nisso tudo com cuidado nota que não. Observando-se com perícia, percebe-se que aquela tabela indica uma operação acadêmica quase diabólica que os entendidos chamam de *arar o mar* e os populares denominam de *retirar leite de pedras*.

Nem sempre *arar o mar* ou *retirar leite de pedras* surte efeito positivo. Mas, no caso presente – para o bem da tese, do livro e do seu autor –, o resultado foi formidável. É para se ter uma dimensão desse êxito basta seguir os resultados nos capítulos quarto e quinto para se apreciar um verdadeiro inferno de dados virando um mar de águas limpas, claras e tranquilas de compreensão de realidades brasileiras.

Ao fundo, esses capítulos quarto e quinto apresentam uma rigorosa *análise fatorial*, marinada em sofisticados arranjos quantitativos, que desemboca no Índice de Desenvolvimento Econômico da Ruralidade (Ider) e no Índice de Desenvolvimento Econômico Local (Idel). Índices – o Ider e o Idel – que propõem, simplesmente, a superação

da abrangência, precisão e qualidade dos resultados aferíveis pelo Índice de Desenvolvimento Humano Municipal (IDH-M) e pelo Índice Firjan de Desenvolvimento Municipal (IFDM-G).

Dito sem meias-palavras: uma contribuição imensa e, em muitos aspectos, acadêmica e universitariamente, sem par. Ou, como se diz no popular, uma contribuição de ouro.

Sim, o Ider e o Idel, feitos como foram feitos, são ouro. Ouro bruto, ouro negro, ouro raro. Um tipo de ouro que amplifica a percepção da complexidade do desenvolvimento econômico brasileiro, oferece uma maior dimensão das desigualdades locais e regionais do país e aumenta a consciência de todos sobre as urgências da sociedade brasileira.

Como não ler esse livro?

Mas o livro ainda tem e diz mais.

No seu primeiro capítulo, vão indicados passeios interessantes pelas controvérsias insuperáveis referentes ao desenvolvimento econômico e, no seu segundo capítulo, vão desenhados esforços decididos para a incorporação da noção de ruralidade na discussão sobre desenvolvimento. Mas é – como já mencionado – do terceiro capítulo em diante que vão contidas as contribuições essenciais do livro.

De toda sorte, quem acompanhar o texto na sequência, de uma capa a outra,

vai perceber ainda mais a força de toda a tessitura do argumento se misturando à exposição feito uma rota em aceleração.

Mas, por claro, não dá para se concordar com tudo que o livro propõe. Existem passagens – especialmente nos capítulos primeiro e segundo – que deveriam desaparecer e outras muitas – notadamente nos capítulos terceiro, quarto e quinto e nas Considerações Finais – que mereceriam ser desdobradas. Sobre a distribuição histórica dos dados enquadrados entre 2004-2006 e 2012-2013, por outro lado, muito mais poderia ser dito. E não restam dúvidas que o autor tem muito mais a dizer.

Avançando para outro horizonte e outros trabalhos, como sugestão, todos os dados do livro poderiam ser atualizados. Como sabido, já saíram as amostras do Censo Agropecuário de 2017 e do Censo Geral de 2020-2022. Da mesma maneira, poderia existir uma maior problematização sobre os possíveis impactos do Ider e do Idel na formulação e reformulação de políticas públicas vigentes. Por fim, a parte mais técnica, os capítulos do terceiro ao quinto, poderia ser transformada em cartilhas direcionadas aos homens práticos tomadores de decisão. Eles precisam desse ouro para se guiar.

REFERÊNCIAS

- ANDERSON IMBERT, E. *Historia de la literatura hispanoamericana*. México, Fondo de Cultura Económica, 1995.
- ARIDA, P. *História do pensamento econômico como teoria e retórica*. Rio de Janeiro, PUC-RJ, 1983.
- BAIROCH, P. *Economics and World History: myths and paradoxes*. Chicago, The University of Chicago Press, 1993.
- BENTLEY, J. H. *The Oxford handbook of world history*. Oxford, Oxford University Press, 2015.
- BIDERMAN, C.; COZAC, L. F. L.; REGO, J. M. *Conversas com economistas brasileiros*. São Paulo, Editora 34, 1997.
- BLOOM, H. *O cânone ocidental*. Trad. Marcos Santarrita. São Paulo, Objetiva, 2001.
- BORGES, J. L. *Obras completas*. Buenos Aires, Emecé, 1974.
- CAMPOS, R. *A lanterna na popa: memórias*. Rio de Janeiro, Topbooks, 1999.
- CAPRA, F. *O tao da física: uma análise dos paralelos entre a física moderna e misticismo oriental*. São Paulo, Cultrix, 1990.
- CARLSNAES, W.; RISSE, T.; SIMMONS, B. *Handbook of international relations*. Londres, Sage, 2002.
- DAMÁSIO, A. *O erro de Descartes*. Trad. Dora Vicente. São Paulo, Companhia das Letras, 2012.
- FRANCO, G. H. B. *A moeda e a lei – Uma história monetária do Brasil, 1933-2013*. Rio de Janeiro, Zahar, 2017.
- FURTADO, C. *A fantasia desfeita*. São Paulo, Paz e Terra, 1989.
- FURTADO, C. *A fantasia organizada*. Rio de Janeiro, Paz e Terra, 1985.
- FURTADO, C. *Essencial*. São Paulo, Companhia das Letras, 2013.
- FURTADO, C. *Formação econômica do Brasil: edição comemorativa, 50 anos*. São Paulo, Companhia das Letras, 2009.
- GADAMER, H.-G. "Hermeneutics and the social sciences". *Cultural Hermeneutics*, vol. 2, 1975, pp. 307-16.
- GAUCHON, P. (coord.). *Le monde – Manuel de géopolitique et de géoéconomie*. Paris, Puf, 2008.
- GOSWAMI, A. *O médico quântico: orientações de um físico para a saúde e a cura*. São Paulo, Cultrix, 2006.
- HARTOG, F. *Régimes d'historicité: presentisme et expérience du temps*. Paris, Seuil, 2003.
- HEGEL, G. W. F. *Raison dans l'histoire – Introduction aux leçons sur la philosophie de l'histoire du monde*. Paris, Seuil, 2011.
- ISAACSON, W. *Einstein*. Trad. Celso Nogueira. São Paulo, Companhia das Letras, 2007.
- KOSELLECK, R. *Futuro passado – Contribuição à semântica dos tempos históricos*. Trad. Wilma Patrícia Mass e Carlos Almeida Pereira. Rio de Janeiro, Contraponto, 2006.
- MORIN, E. *Sociologie de la complexité*. Paris, Agogée, 2010.
- ONFRAY, M. *Decadence. De Jesus à Bin Laden. Vie et mort de l'Occident*. Paris, Flammarion, 2017.
- REGO, J. M. *Revisão da crítica: metodologia e retórica na história do pensamento econômico*. São Paulo, Editora Bial, 1991.

- RENZI, A. et al. "Crescimento econômico nos municípios do Paraná-Brasil: uma análise com econometria espacial". *RPER*, n. 61, 2022, pp. 151-67.
- RENZI, A., HENZ, A. P., & RIPPEL, R. "Desenvolvimento econômico: do crescimento à prosperidade". *Revista Brasileira de Gestão e Desenvolvimento Regional*, n. 15, v. 6, 2019.
- ROS, J. *Rethinking economic development, growth and institutions*. Oxford, Oxford University Press, 2013.
- SANTOS, M. *A natureza do espaço*. São Paulo, Hucitec, 1996.
- SKIDELSKY, R. *John Maynard Keynes. Fighting for Britain, 1937-1946*. Londres, Viking Press, 2000.

A **revistausp** RECOMENDA A SEUS COLABORADORES QUE ENVIEM SEUS TEXTOS DE ACORDO COM AS NORMAS ABAIXO:

1. Texto com, no máximo, 30.000 caracteres.
Textos maiores devem ter a aprovação prévia da redação e do Conselho Editorial. O artigo deve ser enviado por *e-mail* (revisusp@usp.br).
2. O autor deve mandar uma breve nota biobibliográfica que indique onde ensina e/ou pesquisa, sua área de trabalho e principais publicações.
3. Referências bibliográficas devem ser colocadas no texto (sobrenome do autor, ano, página). Os outros dados da publicação devem ser colocados na bibliografia. As notas de rodapé destinam-se a informações ou esclarecimentos adicionais que não podem ser incluídos no texto.
4. Todos os textos devem conter *resumo/abstract* (com, no máximo, mil caracteres) e *palavras-chave/keywords* (com, no máximo, cinco palavras).
5. Textos enviados espontaneamente devem ser inéditos no país. Dados a público previamente, sob forma de palestra, comunicação etc., deve-se informar em nota à parte. Os artigos serão examinados pelo Conselho Editorial, sendo que a **revistausp** não se responsabiliza pela devolução dos não aprovados ou mesmo pela comunicação aos autores.
6. O autor pode enviar/indicar sugestões de ilustração para seu texto. As fotos devem ter alta resolução (300 dpi) e conter legendas e créditos. Imagens retiradas da Internet não serão aceitas.
7. Para receber os exemplares da **revistausp** a que tem direito, o colaborador deve enviar número de telefone, *e-mail* e endereço.

Assine

Assine e receba em casa a publicação de cultura mais completa do país. Você terá a satisfação de ler a cada volume vários textos assinados pelos mais renomados autores em suas áreas. Multidisciplinar e sem preconceitos, a **revistausp** não privilegia esse ou aquele enfoque, esse ou aquele grupo; é aberta a todas as tendências.

Para adquirir uma assinatura anual da **revistausp** (4 edições), basta fazer um depósito em conta (Banco do Brasil - agência 7009-2; c/c 130010-5) em nome da Edusp - Editora da USP, CNPJ 63.025.530/0072-06, no valor de R\$ 70,00 (+ frete). Se você for renovar sua assinatura, ela sai por R\$ 60,00 (+ frete). O preço da revista avulsa é R\$ 20,00 (+ frete). Enviar comprovante de depósito, incluindo o valor do frete, para mariacat@usp.br, juntamente com os dados do assinante (nome/razão social, nome fantasia, CNPJ ou CPF, endereço, Inscrição Estadual/Municipal, telefone, *e-mail* para envio da nota fiscal).

www.usp.br/revistausp
revisusp@usp.br
(11) 3091-4403

Nosso *site* pode ser acessado por celular utilizando-se o QR Code que consta na contracapa da revista.

A **revistausp**, publicação trimestral da Superintendência de Comunicação Social da USP, é editada desde 1989, mantendo sempre a mesma estrutura. A cada número, além da relação dos Dossiês aqui apresentada, temos as seções Textos, Livros e Arte. Abaixo, os números ainda disponíveis.

6 Europa Central	65 20 Anos de Redemocratização	87 Música Brasileira	98 Memória	109 Democracia na América Latina	121 Artes & Letras	132 Bicentenário da Independência: Economia
7 Tecnologias	66 Ano Internacional da Física	88 Humor na Mídia	99 Futebol	110 Ética e Sociedade	122 Feminismos	133 Bicentenário da Independência: Cultura e Sociedade
33 Aids	72 Ar/Fogo	89 Ciência, Tecnologia e Inovação	100 Educação	111 Música Popular Brasileira na USP	123 Histórias Culturais Transatlânticas	134 Bicentenário da Independência: Política
39 Rumos da Universidade	73 Financiamento da Pesquisa no Brasil	90 Marketing Político	101 Justiça Brasileira	112 Americanistas	124 Inteligência Artificial	135 Bicentenário da Independência: Ciência e Tecnologia
47 Alternativas para o Século XXI	76 Pensando o Futuro: Ciências Exatas	91 Catástrofes	102 Metrópoles	113 Amazônia Azul	125 Saramago	136 Integração Latino-Americana
50 Revista Cinquenta	78 Gestão e Política na Universidade Pública	92 Redes Sociais	103 Clima	114 Interculturalidades	126 Semiótica e Cultura	137 Vida Escolar
51 Saúde	80 Bibliotecas Digitais/ Bibliotecas Virtuais	93 Caminhos do Desenvolvimento	104 Energia Elétrica	115 Politicamente Correto	127 Ensino Público	138 Linguística da Vida Contemporânea
54 Os Sertões: Cem Anos	83 Nabuco e a República	94 Semana de Arte Moderna	105 Universidade em Movimento	116 Pós-Verdade e Jornalismo	128 Saúde Pública	139 Edição e Política
55 Revolução Virtual	84 Vinte Anos da Queda do Muro	95 Desafios do Pré-sal	106 Crise Hídrica	117 Copas do Mundo	129 Segurança Pública	140 Literatura de Entretenimento
62 Cosmologia	85 Balanço da Crise Mundial	96 Alcoolismo	107 Saúde Urbana	119 Direitos Humanos	130 Independências Latino-Americanas	
64 Brasil Rural	86 Cibercultura	97 Computação em Nuvem	108 Jogos Olímpicos	120 Religião e Modernidade	131 Pandemia: Valores em Crise	

Você pode acessar todas as edições da **revistausp** no Portal de Revistas da USP:

www.revistas.usp.br/revusp/