

Fábio Luiz Tezini Crocco
Delmo Mattos da Silva
Cassiano Terra Rodrigues
Brutus Abel Fratuze Pimentel
(Organizadores)



**Perspectivas
Éticas
para a IA**
dilemas e desafios



ORGANIZADORES
FÁBIO LUIZ TEZINI CROCCO,
DELMO MATTOS DA SILVA
CASSIANO TERRA RODRIGUES
BRUTUS ABEL FRATUCE PIMENTEL

PERSPECTIVAS ÉTICAS PARA A IA
DILEMAS E DESAFIOS

DOI: 10.29327/5798915

AUTORES

CAROLINA VAITIEKUNAS PIZARRO, DANIELA AMÉRICA DA SILVA,
DELMO MATTOS DA SILVA, JOHNNY CARDOSO MARQUES,
JOSÉ ALEXANDRE MATELLI, LIGIA PAVAN BAPTISTA,
MAURÍCIO CÉSAR DELAMARO, RAFAEL EVANGELISTA,
RODOLFO AVELINO E WALTER TEIXEIRA LIMA JUNIOR

SÃO PAULO
2026



© 2026 by Fábio Luiz Tezini Crocco, Delmo Mattos da Silva, Cassiano Terra Rodrigues e Brutus Abel Fratuze Pimentel

Editoras: Fernanda Aiub e Monica Aiub

Revisão: Monica Aiub

Projeto gráfico e capa: Fernanda Aiub

Imagem de capa: Geralt e *χρῖς* by Pixabay

Conselho Editorial

Edna Alves de Souza (UFAC)

Homero Santiago (USP)

Ivo Assad Ibri (PUC-SP)

João de Fernandes Teixeira (UFSCar)

José Barrientos-Rastrojo (Universidade de Sevilha)

Liana Gottlieb (ECA-USP)

Lucas Nascimento Machado (Universidade de Verona)

Maria Eunice Quilici Gonzalez (UNESP)

Maxwell Morais de Lima Filho (UFCA)

Oswaldo Pessoa Jr. (USP)

Ricardo Lopes Correia (UFRJ)

Sofia Miguens (Universidade do Porto)

Dados Internacionais de Catalogação na Publicação (CIP) (Câmara Brasileira do Livro, SP, Brasil)

Perspectivas éticas para a IA [livro eletrônico] :
dilemas e desafios / organizadores Fábio Luiz
Tezini Crocco, Delmo Mattos da Silva,
Cassiano Terra Rodrigues. -- 1. ed. --
São Paulo : Editora Aiub, 2026.
PDF

Vários autores.

Outros organizadores: Delmo Mattos da Silva,
Cassiano Terra Rodrigues, Brutus Abel Fratuze
Pimentel.

Bibliografia.

ISBN 978-65-988889-7-8

1. Colonialismo 2. Direitos humanos 3. Ética
4. Inteligência artificial 5. Relações humanas -
Recursos de rede de computador 6. Teoria do
conhecimento 7. Transparência - Aspectos sociais
I. Crocco, Fábio Luiz Tezini. II. Silva, Delmo
Mattos da. III. Rodrigues, Cassiano Terra.
IV. Pimentel, Brutus Abel Fratuze.

26-339025.0

CDD-303.483

Índices para catálogo sistemático:

1. Inteligência artificial : Relações humanas :
Sociologia 303.483

Aline Grazielle Benitez - Bibliotecária - CRB-1/3129



Tel.: (11) 93334-0408

E-mail: editoraaiub@gmail.com

www.editoraaiub.com.br

*Só eu posso pensar se deus existe
Só eu
Só eu posso chorar quando estou triste
Só eu
Eu cá com meus botões de carne e osso
Hum, hum
Eu falo e ouço
Hum, hum
Eu penso e posso
Eu posso decidir se vivo ou morro
Porque
Porque sou vivo, vivo pra cachorro
E sei
Que cérebro eletrônico nenhum me dá socorro
Em meu caminho inevitável para a morte*

Gilberto Gil¹

A curto prazo, a chegada de cérebros feitos de silício é um imenso desafio para a maioria das profissões: como existir em um mundo em que a inteligência não será mais contingenciada?

Laurent Alexandre e Jean-Michel Besnier²

¹ GIL, Gilberto. **Cérebro eletrônico**. Fonte: álbum Gilberto Gil, 1969. Philips, LP R 765.087 L, faixa 1, lado A.

² ALEXANDRE, Laurent e BESNIER, Jean-Michel. **Os robôs fazem amor? O transhumanismo em doze questões**. Trad.: Gita K. Guinsburg; prefácio Marta M. Kanashiro. 1ª ed. São Paulo: Perspectiva, 2022, p. 85.

Sumário

PREFÁCIO	
O CHATGPT ESTÁ NOS TORNANDO ESTÚPIDOS?	9
<i>João de Fernandes Teixeira e Cassiano Terra Rodrigues</i>	
APRESENTAÇÃO	15
<i>Delmo Mattos da Silva e Fábio Luiz Tezini Crocco</i>	
1. O CÓDIGO DE ÉTICA DO DESIGNER FRENTE AOS AVANÇOS DA INTELIGÊNCIA ARTIFICIAL: REFLEXÕES E PROPOSIÇÕES	21
<i>Carolina Vaitiekunas Pizarro</i>	
2. RESPONSABILIDADE NA ERA DA IA: TRANSPARÊNCIA E MERCOSUL	51
<i>Daniela América da Silva, Johnny Cardoso Marques, Delmo Mattos da Silva</i>	
3. FORMAÇÃO DOCENTE E IA: A EXPERIÊNCIA DA UNESP	77
<i>José Alexandre Matelli e Maurício César Delamaro</i>	
4. ÉTICA NA INTELIGÊNCIA ARTIFICIAL: DESAFIOS E RECOMENDAÇÕES – UMA PERSPECTIVA DE DIREITOS HUMANOS E IGUALDADE DE GÊNERO	109
<i>Ligia Pavan Baptista</i>	
5. DO COLONIALISMO À SOBERANIA DIGITAL: ÉTICA E JUSTIÇA PARA O FLORESCIMENTO HUMANO	129
<i>Rafael Evangelista e Rodolfo Avelino</i>	
6. SIMBIOSE HUMANO-MÁQUINA COMPUTACIONAL: O LEGADO TECNOLÓGICO VISIONÁRIO DE LICKLIDER	147
<i>Walter Teixeira Lima Junior</i>	
QUEM SOMOS?	161

PREFÁCIO

O CHATGPT ESTÁ NOS TORNANDO ESTÚPIDOS?

DOI: 10.29327/5798915.1-1

É mesmo um desafio escrever um prefácio para este *Perspectivas Éticas para a IA: dilemas e desafios*. Além da diversidade de autores e capítulos, não é tarefa de pouca monta apresentar reflexões que tentam decifrar o presente. É ainda mais quando as mudanças do presente são rápidas e profundas. Por isso, assim como o próprio livro, seu prólogo só pode ser um convite à reflexão. Dada a amplitude do tema, o que buscamos fazer aqui é apenas esboçar algumas linhas mestras para que cada leitor ou leitora possa, a partir disso, alinhar as próprias inferências suscitadas pela leitura.

Temos de reconhecer: os desenvolvimentos da tecnologia conhecida como Inteligência Artificial generativa (IA, daqui em diante) ainda não dão sinais de esgotamento. De fato, o “boom da IA”, se essa expressão for correta, trouxe consequências consideradas inéditas na história, ao menos até o ponto em que é possível julgar no calor do momento. Da inédita acessibilidade da tecnologia – praticamente qualquer pessoa com acesso à internet, hoje, já interage com complexos modelos de linguagem, conscientemente ou não – aos imensos impactos econômicos e sociais, a IA tornou-se praticamente onipresente na vida social, permeando toda a cadeia produtiva, da gestão da rede de abastecimento à detecção de fraudes no sistema financeiro, sem falar no impacto sobre os empregos e a necessidade de regulamentação jurídica para a privacidade e segurança dos usuários. Não por menos, sentimos que o futuro pertence à IA.

E se é assim, mais importante do que saber o que faremos com a IA, é refletir criticamente – sem medo, mas tampouco com ingenuidade – sobre o que a IA pode fazer conosco. Pois é bem verdade que não podemos aceitar tecnologias simplesmente por serem inovadoras, o que só aumenta a nossa responsabilidade política e social. Essa é a

questão central deste livro, em torno da qual gravitam todas as reflexões aqui publicadas.

Um dos riscos que têm sido apontados com o uso crescente da IA é a atrofia da cognição humana. A IA nos desincumbe do aprendizado e do exercício de diversas atividades mentais que passam a ser exercidas por algoritmos automáticos. A presença desses algoritmos tem aumentado desde a invenção dos LLMs ou *Large Language Models*, como o ChatGPT. Há um temor de que essa atrofia se manifeste de forma acentuada nas próximas gerações, que perderiam a capacidade de aprender e criar, sobretudo na área da linguagem. Uma padronização crescente da comunicação é um sintoma dessa tendência. Obras como *A Guerra das Inteligências na era do ChatGPT*, de Alexandre (2024), têm, progressivamente, apontado para esse problema.

O paradoxo desta situação é que, se projetada para as próximas gerações, ela terá como consequência uma incapacidade crescente de programar máquinas superinteligentes. Como no mito do Ouroboros, a IA devoraria sua própria cauda.

No entanto, essa afirmação deve ser nuançada. Por acaso as crianças perderam sua capacidade de aprender aritmética por causa do uso de calculadoras? Uma calculadora é bem diferente do cérebro de uma criança. Se ambos resolvem problemas aritméticos, fazem-no por caminhos muito diferentes. Podemos, então, distinguir entre inteligência natural e inteligência artificial ou maquínica. Professores de filosofia imediatamente se lembrarão da dicotomia instaurada pela filosofia moderna entre a natureza e a cultura, e o quanto a contemporaneidade se esforçou para superar a rigidez desse dualismo. Então, se já é uma realidade a existência de máquinas inteligentes que se autorregulam, em certa medida e para certas tarefas, algo que até bem pouco tempo atrás era considerado uma capacidade exclusivamente humana, ou ao menos orgânica (na falta de termo melhor), não é de estranhar a proliferação de discursos sensacionalistas sobre as capacidades da IA e o espanto do público em geral com as novidades tecnológicas. Se estava certo Aristóteles e o espanto é mesmo o estado que leva à filosofia, teremos ainda muito a filosofar.

A distinção mencionada entre o natural e o artificial, por mais esquemática que seja, fornece um caminho para pensar a questão que enunciamos. A IA não entrará, necessariamente, em decadência num futuro próximo por causa de nossa incapacidade de programar máquinas complexas. Num futuro próximo, parece bastante seguro afirmar que ela não precisará mais incluir seres humanos na produção de algoritmos inteligentes. As máquinas superinteligentes serão capazes de se autoprogramar. A crescente autonomia das máquinas, sua capacidade de autoprogramação e as possibilidades abertas pelo aprendizado profundo estão eliminando, progressivamente, a participação de seres humanos numa escalada planetária da inteligência. Isso responde à dúvida acerca de quem ou o que programará máquinas superinteligentes no futuro. Essa será uma tarefa para a inteligência maquínica e não para a inteligência natural.

Os maiores entusiastas dessa visão são os transhumanistas. Alguns deles, especialmente Ray Kurzweil (2024), defendem o advento da singularidade, no qual as máquinas serão capazes de produzir outras máquinas, levando a uma explosão de inteligência. Nesse processo, os humanos serão marginalizados e a inteligência progredirá sem eles. A inteligência maquínica se tornará totalmente independente da inteligência natural. Podemos, neste caso, falar de uma bifurcação entre inteligência natural e inteligência artificial.

Não há muitos argumentos para acreditarmos na singularidade a não ser a observação de tendências presentes. Os resultados têm sido promissores, mas sabemos que nem sempre os raciocínios indutivos são corretos. Cabral temia navegar abaixo do Equador, pois projetava que o calor aumentaria tanto que incendiaria suas embarcações. Como raciocínio indutivo, a hipótese de Cabral estava certa, mas sua conclusão era falsa. Já o filósofo estadunidense Daniel Dennett, recentemente falecido, afirmou que a singularidade não passaria de uma lenda urbana (Dennett, 2015). Uma história linear, sem catástrofes, é o pressuposto não demonstrado dos transhumanistas.

É possível afirmar que há um erro categorial nas discussões sobre o futuro da inteligência. As superinteligências, ou inteligências artificiais, não estão ligadas a objetivos e recompensas como ocorre com a inteligência natural. Deep Blue não se sentiu satisfeito por ter conseguido derrotar Kasparov numa partida de xadrez decisiva. Tampouco a máquina que derrotou Lee Sedol, o campeão mundial de Go, se sentiu feliz. Esses exemplos indicam que a inteligência natural nunca mais alcançará uma superinteligência que ela tenha produzido. A inteligência maquínica, por sua vez, precisará cada vez menos da inteligência humana para evoluir.

O temor de que nos tenhamos tornado incapazes de produzir tecnologia é infundado, pois estamos usando a palavra inteligência para nos referirmos a fenômenos diferentes: de um lado, a inteligência natural, ligada a desejos, objetivos e satisfação, e, de outro, as inteligências artificiais que podem progredir independentemente desses fatores. Uma inteligência se torna artificial quando não precisa mais da mente humana para se desenvolver e ultrapassa a mimetização das operações cognitivas do homem. O desafio, daqui para a frente, será alinhar esses dois tipos de inteligência nas nossas sociedades, tendo como pressuposto evitar essa confusão categorial. Políticas tecnológicas mostram-se cada vez mais necessárias num mundo em que as nossas máquinas já exibem suficiente autonomia para não nos considerarmos capazes de controlá-las.

Resta saber se é possível sustentar que o ChatGPT não inibe o raciocínio crítico. Supomos que, ao contrário, ele pode levar à sua expansão, ao auxiliar as técnicas de redação. É preciso saber se os sequenciamentos de palavras que ele oferece inibem o pensamento ou se servem como estímulo para a criação, na medida em que incitam o reconhecimento de novas possibilidades na escrita. Como neste caso não se trata de simular operações matemáticas ou aritméticas, o ChatGPT pode funcionar como produtor de associações livres ao estimular a memória e, com ela, a criatividade. Mas esta é uma visão polêmica do ChatGPT e, por isso, sujeita a muitos ataques.

É possível que tudo o que tenhamos afirmado até aqui se baseie em falsas imagens e pressupostos que temos acerca de nós mesmos. Como assinalou George Orwell, cada geração se imagina mais inteligente do que a que a precedeu e mais sábia do que a que vem depois dela. Isso é uma ilusão, e deve-se reconhecê-la. Temos de manter nossa própria visão de mundo, mesmo ao preço de ela parecer antiquada, pois essa visão de mundo brota de experiências que a geração mais jovem não teve. É fundamental, portanto, resistirmos à equivocada ideia de que a transmissão da experiência social pode ser automatizada. De fato, o conhecimento é uma construção coletiva e não necessariamente obedece à lógica do consumo de mercadorias. Mesmo que no mundo atual as pessoas se relacionem em massa com ciência e tecnologia comprando produtos – e os LLM’s são produtos desenvolvidos por grandes empresas capitalistas, não podemos esquecer – essa relação não é necessária e há quem defenda que sequer é desejável.

Seja como for, é um fato social que a imensa maioria da população identifique invenção e descoberta científico-tecnológica com inovação produtiva. Talvez isso explique certos tipos de saudosismo ou de aversão à tecnologia, comumente ligados à sensação de uma certa queda na qualidade de nossos utensílios tecnológicos, atribuída a uma decadência da qualidade do trabalho. Mas isso pouco tem que ver com a IA. Tal sensação reflete, sim, o desinteresse econômico em investir em itens tecnológicos cuja produção já se encontra consolidada e que frequentemente é realizada por proletariados remotos que oferecem trabalho barato. Daí uma percepção difusa de que geladeiras, máquinas de lavar, computadores e outros itens de tecnologia básica vêm piorando. É equivocado atribuir esse fenômeno à IA ou a possíveis desgastes intergeracionais.

Evidentemente, se recordarmos o que dizia Marx sobre o objetivo do desenvolvimento da maquinaria como forma de organização produtiva, é inegável que a IA atualiza o sonho capitalista de automação completa da produção, possibilitando um controle muito maior sobre o trabalho humano,

o único capaz, na teoria marxista, de gerar valor novo. E realmente podemos identificar na reconfiguração do trabalho muito mais do que uma tendência, antes, uma realidade inescapável: ao mesmo tempo que a IA parece reatualizar certos aspectos já descritos por Marx, como a intensificação do trabalho e o aumento do número de desempregados e excluídos do mercado, ela também mostra seus limites quando certas habilidades específicas e que não podem ser mecanizadas passam a ser valorizadas. Os limites para a automação são dados pelo conhecimento humano acumulado por eras de tentativa e erro. Como uma máquina poderia ensinar alguém a nadar ou andar de bicicleta? Pode uma máquina aprender a reconhecer odores e sabores? Ou ainda, como ensinar uma máquina a relacionar certos odores e sabores a experiências de prazer ou de dor?

Não ousamos oferecer respostas a tais questionamentos. Se aqui os registramos, é para dar a público um gosto do que é possível encontrar neste livro e, assim, reiterar o convite à reflexão. Esperamos que a leitura dos capítulos seja um convite, melhor dizendo, à imaginação viva, pela qual consigamos sonhar com um mundo no qual a inteligência se estenda a caminhos mais criativos, mais diversos e mais justos do que os que nos trouxeram até aqui. Afinal, se criamos até máquinas que aprendem, por que nos recusaríamos a aprender com as nossas próprias criações?

João de Fernandes Teixeira e Cassiano Terra Rodrigues

REFERÊNCIAS

- ALEXANDRE, Laurent. *A Guerra das Inteligências na era do ChatGPT*. Trad. Idalina Lopes. 1ª ed. Barueri, SP: Amarylis, 2024. (publicado originalmente em 2023).
- DENNETT, Daniel. The Singularity – an Urban Legend? In: *Edge.org*, seção “Annual question”, 2015. Disponível em: <https://www.edge.org/response-detail/26035>. Acesso em 24/01/2026.
- KURZWEIL, Ray. *A Singularidade está mais próxima: A fusão do ser humano com o poder da Inteligência Artificial*. Trad. Renato Marques. São Paulo: Goya, 2024.

APRESENTAÇÃO

DOI: 10.29327/5798915.1-2

A ética na inteligência artificial (IA) tornou-se um dos temas centrais na atualidade, não apenas no campo tecnológico, mas também na filosofia, no direito, nas ciências sociais, no jornalismo, nas relações internacionais e nas engenharias. Trata-se de um campo interdisciplinar que busca compreender como projetar, usar, regular e avaliar sistemas de IA de maneira responsável, garantindo que essas tecnologias sirvam ao bem-estar humano e não produzam danos indevidos.

Luciano Floridi, especialmente em *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities* (2023), menciona que a ética da IA não deve ser compreendida apenas como um conjunto de regras aplicadas a artefatos técnicos, mas como parte de uma ética da informação mais ampla, voltada à preservação e ao florescimento do ambiente informacional (*infosfera*). Segundo Floridi (2023), os sistemas de IA são agentes morais no sentido funcional, ainda que não possuam intencionalidade ou consciência, porque exercem impactos reais, sistemáticos e em larga escala sobre indivíduos, instituições e estruturas sociais.

Nesse contexto, o desafio ético central reside em “reconfigurar as responsabilidades humanas” ao longo de todo o ciclo de vida dos sistemas de IA: concepção, desenvolvimento, treinamento, implementação, uso e governança. Floridi destaca que a pergunta ética fundamental não é apenas “o que a IA pode fazer?”, mas “o que devemos permitir que a IA faça” em um ambiente social justo e sustentável (Floridi, 2023).

Por outro lado, Mark Coeckelbergh, em *AI Ethics* (2020), sustenta que a ética da inteligência artificial deve ser compreendida menos como uma indagação acerca das supostas capacidades morais das máquinas e mais como uma investigação crítica sobre as transformações das práticas humanas mediadas por tecnologias inteligentes. Nessa perspectiva, a questão central não consiste em saber se a IA pode ser

considerada um agente moral, mas em analisar de que modo esses sistemas reconfiguram relações sociais, expectativas normativas e estruturas de poder.

A partir de uma abordagem relacional, Coeckelbergh argumenta que a moralidade não reside em propriedades internas dos artefatos técnicos, mas emerge das interações entre seres humanos, tecnologias e contextos institucionais específicos. Ainda que desprovidos de consciência ou intencionalidade, os sistemas de IA exercem influência significativa sobre processos decisórios, formas de responsabilização e dinâmicas sociais, tornando indispensável uma análise ética que vá além do artefato isolado e considere os ecossistemas sociotécnicos nos quais essas tecnologias são desenvolvidas e utilizadas.

A partir dessa perspectiva, princípios como beneficência, não maleficência, justiça, explicabilidade e responsabilidade não devem ser tratados como adições externas ao projeto técnico, mas como “critérios constitutivos do próprio design dos sistemas” (Russell, 2019). A ética da IA, portanto, deixa de ser reativa, ou seja, acionada apenas após danos ou controvérsias, para assumir um caráter proativo e estrutural, integrando-se às práticas de engenharia, às políticas públicas e às decisões organizacionais.

Fruto dos esforços teóricos do Grupo de Pesquisa IDEIA – Inovação e Desafios Éticos da IA (CNPq), das discussões promovidas no I Seminário de Ética e Inteligência Artificial do Instituto Tecnológico de Aeronáutica – ITA (2025) e do apoio da ITAEx (Ex-alunos Apoiando o ITA), o presente livro reúne diferentes perspectivas sobre a ética na inteligência artificial (IA), abordando seus principais dilemas sociais, tensões normativas e desafios práticos a partir de múltiplos campos disciplinares. Cada autor(a) contribui para o debate a partir de sua área de atuação, oferecendo uma leitura plural sobre os impactos éticos, sociais, políticos e técnicos dos sistemas de IA.

No primeiro capítulo, intitulado *O código de ética do designer frente aos avanços da Inteligência Artificial: reflexões e proposições*, a pesquisadora Carolina Vaitiekunas Pizarro dis-

cute o papel do designer na era da IA, analisando como os princípios éticos da profissão podem e devem orientar o desenvolvimento de soluções que incorporam IA. O capítulo estabelece um diálogo entre o Código de Ética do Designer (*World Design Organization*) e as diretrizes da UNESCO, propondo atualizações necessárias para garantir a proteção do usuário.

Na sequência, em *Responsabilidade na era da IA: transparência e Mercosul*, os autores Daniela América da Silva, Johnny Cardoso Marques e Delmo Mattos da Silva exploram o complexo dilema da responsabilidade quando decisões são delegadas a sistemas de IA. Com foco na transparência como mecanismo mitigador, o texto analisa como os países do Mercosul têm abordado a explicabilidade algorítmica e propõe uma visão integrada para a regulamentação regional.

O terceiro capítulo, escrito por José Alexandre Matteli e Maurício César Delamaro e intitulado *Formação docente e IA: a experiência da Unesp*, relata a experiência da Universidade Estadual Paulista (UNESP) na preparação de seu corpo docente para os desafios pedagógicos impostos pela IA generativa. Destaca-se a importância de uma formação pedagógica sólida, que oriente e permita aos professores utilizarem a IA de modo intencional e ético, com base em políticas institucionais bem fundamentadas.

No capítulo *Ética na Inteligência Artificial: desafios e recomendações – uma perspectiva de direitos humanos e igualdade de gênero*, Ligia Pavan Baptista aborda os riscos éticos da IA sob a ótica dos direitos humanos e da igualdade de gênero, discutindo temas como discriminação algorítmica, privacidade, desinformação e ameaças à democracia. O capítulo apresenta diversas experiências institucionais e acadêmicas nacionais e internacionais, iniciativas e reflexões com base nas recomendações da UNESCO e defende a necessidade de regulamentações que assegurem um uso justo, sustentável e centrado no ser humano, em alinhamento com os Objetivos de Desenvolvimento Sustentável da ONU.

O capítulo 5, *Do colonialismo à soberania digital: ética e justiça para o florescimento humano*, de Rafael Evange-

lista e Rodolfo Avelino, parte de uma perspectiva situada no Sul Global para criticar o colonialismo e a colonialidade digitais e defender a autonomia e a justiça para sociedades historicamente minorizadas. O texto questiona perspectivas superficiais e limitadas de utilização do conceito de ética, comumente defendidas por setores empresariais e políticos, e propõe a utilização do termo soberania digital como reivindicação política. Segundo os autores, soberania digital constitui uma categoria analítica mais rigorosa e produtiva, pois está situada socialmente em processos que marcam a tecnopolítica em territórios subalternizados pelo Norte Global e articulada a noções de justiça de dados, florescimento humano e tecnodiversidade.

Por fim, no capítulo 6, *Simbiose Humano-Máquina Computacional: o legado tecnológico visionário de Licklider*, Walter Teixeira Lima Junior recupera a visão pioneira de J.C.R. Licklider, que propôs, já na década de 1960, as bases da computação pessoal, da internet e da relação simbiótica entre humanos e máquinas. A partir da analogia com a simbiose biológica, as ideias seminais de Licklider estabeleceram um paradigma que transformou a forma como interagimos com a tecnologia e nos beneficiamos dela. A discussão ressalta a atualidade desse paradigma para a Inteligência Artificial contemporânea, reforçando que a tecnologia deve ampliar – e não substituir – as capacidades humanas.

Em conjunto, os capítulos sustentam a tese de que a ética da IA não pode ser compreendida como um apêndice normativo ou um conjunto de boas intenções declarativas, mas como uma dimensão constitutiva da própria racionalidade sociotécnica. Ao articular teoria e prática, este livro propõe deslocar o debate da pergunta sobre os limites técnicos da IA para uma reflexão mais profunda sobre quais formas de vida, desenvolvimentos, relações sociais e modelos de justiça estamos dispostos a promover em uma sociedade cada vez mais mediada por sistemas inteligentes.

Nesse horizonte, os dilemas e desafios da ética na inteligência artificial não se configuram como problemas isolados ou meramente técnicos, mas como tensões estruturais

que atravessam os “ecossistemas sociotécnicos contemporâneos”, envolvendo questões como a tensão entre eficiência e justiça, a persistência de vieses e discriminações algorítmicas, a opacidade dos sistemas de aprendizado de máquina e as dificuldades de atribuição de responsabilidade em contextos de decisão automatizada (Floridi, 2023).

Nesse sentido, a concentração política e econômica de poder informacional, a assimetria no acesso a dados e infraestruturas tecnológicas e os desafios da governança global e regional ampliam essas preocupações, ao passo que a crescente delegação cognitiva e a integração entre humanos e sistemas inteligentes exigem a revisão de noções como autonomia, controle humano significativo e *accountability* (Floridi, 2023; Coeckelbergh, 2020). Diante desse cenário, a ética da IA não pode ser reduzida a códigos de conduta ou princípios abstratos, mas deve ser compreendida como um campo de tensão tecnopolítico e uma dimensão constitutiva da “racionalidade sociotécnica contemporânea”, orientada à deliberação democrática, à justiça social e ao florescimento humano.

Delmo Mattos da Silva e Fábio Luiz Tezini Crocco

REFERÊNCIAS

COECKELBERGH, Mark. *AI Ethics*. Cambridge, MA: MIT Press, 2020.

FLORIDI, Luciano. *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities*. Oxford: Oxford University Press, 2023.

RUSSELL, Stuart. *Human Compatible: Artificial Intelligence and the Problem of Control*. New York: Viking, 2019.

**O CÓDIGO DE ÉTICA DO DESIGNER FRENTE AOS
AVANÇOS DA INTELIGÊNCIA ARTIFICIAL:
REFLEXÕES E PROPOSIÇÕES**

**THE DESIGNER'S CODE OF ETHICS AND
ADVANCEMENTS IN ARTIFICIAL INTELLIGENCE:
REFLECTIONS AND PROPOSALS**

DOI: 10.29327/5798915.1-3

*Carolina Vaitiekunas Pizarro*³

RESUMO: O campo do design atualmente pode não apenas integrar a IA como parte de seus processos de criação, como também pode atuar projetando soluções que contenham a IA. No primeiro caso, a IA se apresenta como recurso para otimizar tarefas simples da rotina projetual, mas no segundo, a IA é a tecnologia integrante ou de base da solução projetada. Em ambos os casos a postura ética por parte do profissional designer é essencial, mas especificamente no segundo, ela passa a ter um papel definidor na concepção da solução visando a salvaguarda do usuário. Este artigo busca, a partir de pesquisa bibliográfica associada à análise documental, traçar paralelos entre o Código de ética do designer disponibilizado pela *World Design Organization* e as diretrizes para um uso ético da IA preconizados pela UNESCO. O objetivo é promover reflexões sobre quais pontos do código de ética profissional são atemporais e já atendem às necessidades de um cenário com a presença da IA, e quais precisam ser atualizados para se adequar às demandas oriundas deste cenário.

PALAVRAS-CHAVE: Ética, Design, Inteligência Artificial, Práxis, Direitos humanos

³ Universidade Estadual Paulista (UNESP), Faculdade de Arquitetura, Artes, Comunicação e Design, Bauru, SP, Brasil, carolina.pizarro@unesp.br.

ABSTRACT: The field of design currently can not only integrate AI as part of its creative processes, but can also act by designing solutions that contain AI. In the first case, AI presents itself as a resource to optimize simple tasks in the design routine, but in the second, AI is the integral or foundational technology of the designed solution. In both cases, an ethical stance on the part of the design professional is essential, but specifically in the second, it takes on a defining role in the solution's conception, aiming to safeguard the user. This article seeks, through bibliographic research combined with documentary analysis, to draw parallels between the Designer's Code of Ethics provided by the World Design Organization and the guidelines for ethical AI use advocated by UNESCO. The objective is to promote reflections on which points of the professional code of ethics are timeless and already meet the needs of a scenario with the presence of AI, and which need to be updated to adapt to the demands arising from this scenario.

KEYWORDS: Ethics, Design, Artificial Intelligence, Praxis, Human Rights

INTRODUÇÃO

O avanço da Inteligência Artificial (IA) tem exigido, de diferentes áreas, adequações em suas ações profissionais para lidar com esta tecnologia quando a sua integração faz sentido para um projeto. No caso do campo do design, ele não apenas pode integrar a IA como parte de seus processos de criação e etapas de projeto, como também pode atuar projetando soluções que contenham a IA. No primeiro caso, a IA se apresenta como recurso para otimizar tarefas simples da rotina projetual, mas no segundo, a IA é a tecnologia integrante ou de base da solução projetada. Em ambos os casos a postura ética por parte do profissional designer é essencial, mas especificamente no segundo ela passa a ter um papel crucial na concepção da solução – do ponto de vista de adequações – visando a salvaguarda de direitos do usuário.

O cenário que se apresenta para o campo do design na era da IA adiciona uma camada extra de complexidade para

uma profissão que dá forma à cultura material, a partir de fatores de ordem social, cultural, tecnológica, econômica e política. A escalada da IA impacta e acelera cada uma dessas dimensões (Suleyman; Bhaskar, 2023) trazendo benefícios, mas também riscos éticos (Floridi, 2024) que podem ampliar cenários danosos como os que levam à manipulação, à desinformação, ao totalitarismo (Coeckelbergh, 2020, 2024; Crawford, 2021), e ao preconceito (Noble, 2018; Silva, 2022) entre outros, os quais se tornam desafios para governos e sociedade, mas que também se colocam como desafios éticos para quem projeta soluções contendo IA.

No limite, considerando que o design é o campo responsável por dar forma às soluções que tragam benefícios aos seres vivos, os riscos atrelados à IA se tornam um problema de projeto para a área do design, especificamente na busca por zelar pela salvaguarda de direitos das pessoas que tiverem contato com a solução projetada contendo a tecnologia. É possível, portanto, sintetizar que problemas éticos da IA se tornarão invariavelmente problemas éticos para o design e esta conjuntura se dá diante de um cenário dominado pelas *Big Techs*, no qual o colonialismo de dados é uma realidade (Coudry; Mejias, 2019). Estabelecem-se, assim, dinâmicas de poder sobre as quais o design enquanto área do conhecimento não tem controle e, nesse sentido, cabe aqui a reflexão do papel do design (e do designer) diante desta inquietação que atinge a sociedade na contemporaneidade.

Um dos caminhos possíveis de ação política do design reside nos enfrentamentos éticos que o designer precisa fazer dentro de sua práxis e, para tal, o código de ética da profissão se constitui como um guia de conduta que orienta os profissionais a agir de maneira ética.

Tendo no horizonte esses dilemas, o presente capítulo tem por objetivo traçar um paralelo entre o código de ética do profissional publicado pela *World Design Organization* (WDO) e os princípios elencados nas Recomendações sobre a ética da Inteligência Artificial da UNESCO. A escolha pelo código de ética da WDO se deu pelo fato de esta ser uma organização glo-

bal que pauta mundialmente discussões no âmbito do Design, do mesmo modo a escolha pelo documento da UNESCO se deu por este organismo internacional operar também em escala global e com vários países membros que adotam suas diretrizes, inclusive para a formulação de políticas públicas.

A intenção é identificar, a partir dos princípios para um uso ético da IA preconizados pela UNESCO, quais são os pontos que já podem ser considerados como contemplados pelo código de ética global dos designers adotado pela WDO e refletir sobre quais ainda não o são, de modo a dar visibilidade às áreas em que o código de ética dos designers precisa ser revisto e ampliado.

O DESIGN ENQUANTO AGENTE ÉTICO NA SOCIEDADE IMPACTADA PELA IA

A corrida pela inteligência artificial é uma realidade global. Os sucessivos lançamentos de novos modelos, mais potentes e cada vez mais otimizados, têm moldado decisões econômicas, políticas e jurídicas em diversos países, transformando a sociedade.

Essa complexidade coloca ao designer a necessidade de uma ação cada vez mais consciente no mundo, dada a responsabilidade atrelada ao seu fazer diante de uma tecnologia que avança em uma velocidade maior do que qualquer outra que a precedeu. Diante dos desafios do fazer dos designers para a sociedade, cabe a reflexão de Rodrigues (2009) sobre a necessidade destes profissionais produzirem a conscientização. Apoiando-se em Freire (1980) e sua abordagem sobre a consciência, a autora destaca:

Para ele não basta ter consciência, mas produzir conscientização, e isto significa ultrapassar a visão espontânea da realidade e exercer um conhecimento crítico desta mesma realidade, o que exige uma posição e uma ação. “A conscientização não pode existir fora da ‘práxis’, ou melhor, sem ação – reflexão. Esta unidade dialética constitui, de maneira permanente, o modo de ser ou de transformar o mundo que caracteriza os homens”. Substancializar para o designer significa, portanto, realizar

um trabalho consistente. As atividades projetuais são ações culturais que exigem um “fazer consciente”, desde a escolha de seu objeto de estudo, sua interpretação e sua transmutação (Rodrigues, 2009, p. 3).

Assim, neste artigo, a ação do designer não será tratada somente como prática (execução de uma atividade ou habilidade), mas como práxis, ou seja, esta indissociabilidade entre teoria e prática, de maneira dialética e crítica, que se fará cada vez mais necessária frente ao avanço da IA e seus impactos sociais, sejam estes positivos ou negativos. Em um contexto no qual dinâmicas de poder se instalam também por meio da tecnologia, os designers não podem ser neutros, acríticos ou apolíticos. Antes, podem e devem refletir e mobilizar teoria e prática da profissão, com um olhar permanentemente crítico sobre o cenário que o cerca e para o qual produz, de modo que sua conduta seja ética e emancipatória para os sujeitos nela implicados. Nesse sentido, o código de ética é um caminho para que a profissão seja orientada à práxis consciente e transformadora.

O avanço da IA empurra constantemente os limites da área de design, a qual, segundo Denis (2013), por ser uma área

[...] voltada historicamente para o planejamento de interfaces e para a otimização de interstícios, tende a se ampliar à medida que o sistema se torna mais complexo e à medida que aumenta, por conseguinte, o número de instâncias de inter-relação entre suas partes (Denis, 2013, p. 234).

Este cenário coloca para o design o desafio de identificar como é possível colaborar para que esses avanços sejam benéficos, mas, principalmente, como mitigar ou neutralizar as externalidades negativas ou não previsíveis da tecnologia, de modo a favorecer a sociedade. Sobre o papel do design na missão de mudar a sociedade para melhor, Bonsiepe (2022) reflete:

A difícil questão de saber se o design pode mudar a sociedade levanta a preocupação de estarmos indo além do escopo do que o design pode realmente alcançar. A resposta depende do que entendemos por mudança e das demandas que surgem a partir dessa compreensão. Uma

resposta provável, ainda que rápida: a sociedade muda o design, em vez do contrário. O design, de fato, influencia profundamente e transforma a vida cotidiana na sociedade, mas a profundidade e o alcance dessas mudanças continuam sendo objeto de debate (Bonsiepe, 2022, p. 307, tradução nossa).

A profundidade e o alcance das mudanças provocadas pelo design também são abordados por Fry (2004), para quem se não forem assumidas as devidas responsabilidades éticas, o design pode ser um agente involuntário de *desfuturização*. O autor (Fry, 2020) cunha este termo levando em consideração as consequências do ponto de vista da insustentabilidade ambiental planetária, que o design de certa maneira colaborou involuntariamente para impactar, a partir de sua produção, que foi inserida dentro da lógica da sociedade de consumo:

Isso significa que, no cerne do processo de desenvolvimento da construção dos futuros modernos, houve um contramovimento que assumiu, metaforicamente, vida própria, sendo considerado uma lei, uma lógica e uma racionalidade. Esse contramovimento gerou uma grande contradição: criou um padrão de desenvolvimento que estabeleceu condições de sustentação de curto prazo que minaram a possibilidade de meios de sustentabilidade de longo prazo. A negação do futuro, o desfuturizado, é, portanto, algo que acontece conosco e com o mundo ao nosso redor; estamos dentro e somos parte dele (Fry, 2020, p. 238, tradução nossa).

Entendendo que o design pode ser instrumentalizado dentro de dinâmicas de poder de modo a, involuntariamente, contribuir para cenários problemáticos, deve-se assumir a responsabilidade diante dessa possibilidade quando a IA entra em jogo.

Para Norman (2023), o design está no centro dos problemas que trouxeram a humanidade até aqui, então uma redefinição e reestruturação do design poderia contribuir para encontrar novos caminhos para solucioná-los. Conforme o autor, a profissão

[...] se desenvolveu como uma ferramenta do capitalismo moderno para ajudar a vender os produtos da

Revolução Industrial. Como resultado, o design e a educação em design hoje são impulsionados pela necessidade de que as empresas e os clientes que contratam designers tenham lucro (Norman, 2023, p. 8).

Deve-se ter em conta, portanto, que na era da IA, o campo do design e a atuação dos designers estão sob o mesmo risco que já estiveram ao longo da história da área: tornar-se ferramenta para ajudar a vender produtos, outrora industriais, e agora também digitais, mas que com a ascensão da IA performam oportunidades e riscos sem precedentes (Suleyman; Bhaskar, 2023).

Tratando das dimensões éticas no design, Fry (2004) enfatiza que os designers não têm condições de prever o futuro antecipando as consequências de suas ações antes que elas ocorram, mas podem se propor a pensar, modelar e projetar as consequências do que projetam, e para isso

[...] precisam perguntar e extrapolar: “o que aquilo que eu projeto provavelmente trará à existência e com quais resultados prováveis?”. Tentar responder a esse tipo de pergunta é uma das tarefas cruciais para criar a possibilidade de uma ética do design pensante (Fry, 2004, p. 150, tradução nossa).

Há que se ter clareza que, para que benefícios superem riscos, os desenvolvimentos pautados pela ética são essenciais. Dessa maneira, Floridi (2024), pondera que considerando os avanços no campo da IA, fica evidente que a conformidade com políticas e leis é essencial, mas insuficiente para conduzir a sociedade na direção certa, uma vez que a regulamentação digital é capaz de indicar o que é legal e ilegal, mas não diz quais seriam as melhores escolhas dentro das opções legais para ter uma sociedade melhor “[...] essa é a tarefa da ética digital do lado dos valores e preferências morais, e da boa governança digital do lado da gestão” (Floridi, 2024, p. 143). Para o autor, garantir resultados socialmente preferíveis significa resolver a tensão entre incorporar os benefícios e diminuir os possíveis danos, uma missão que também envolve o escopo de ação do design e do designer. Em suas palavras, esta era da IA é mais do que qualquer outra, a era do design e justifica:

Como o digital está diminuindo as restrições e aumentando as *affordances* à nossa disposição, ele nos oferece uma liberdade imensa e crescente para arrumar e organizar o mundo de várias maneiras a fim de resolver uma variedade de problemas novos e antigos. Obviamente, qualquer design requer um projeto. E, no nosso caso, o que ainda nos falta é um projeto humano para a era digital. Mas não devemos permitir que o *cleaving power* do digital molde o mundo sem um plano. Devemos fazer de tudo para decidir a direção na qual queremos explorá-lo, a fim de garantir que as sociedades da informação que estamos construindo graças a ele sejam abertas, tolerantes, equitativas, justas e favoráveis ao meio ambiente, bem como à dignidade e à prosperidade humanas. A consequência mais importante do *cleaving power* do digital deveria ser um design melhor do mundo (Floridi, 2024, p. 53-54).

O *cleaving power* é o termo empregado pelo autor para referir-se à possibilidade de conectar, desconectar e reconectar que o digital permite, transformando radicalmente a realidade. Se o design pode auxiliar a encontrar uma direção benéfica para o que significa o avanço da IA na sociedade, isso implica na responsabilidade do designer frente ao que compete a ele dentro dos processos e tomadas de decisão em cada projeto, sendo, portanto, um agente que consiga absorver e aplicar a tecnologia de maneira ética, de modo que a solução resultante seja adequada, satisfatória e emancipatória.

Chan (2018) destaca que o design é essencialmente um processo ético, uma vez que escolhas e tomadas de decisão são feitas a todo momento já que os

[...] problemas que os designers escolhem resolver – e porquê – e quem incluir ou excluir como beneficiários deste design não só presumem preferências de escolha, mas também, mais fundamentalmente, posições de valor sobre a vida boa ou que vale a pena (Chan, 2018, p. 184, tradução nossa).

Cabe destacar, contudo, que o design é, por definição, uma área interdisciplinar. Assim, não é uma atividade realizada por um único indivíduo; tipicamente, é um esforço coletivo

e conforme Devon e Van de Poel (2004), lida com as questões que permeiam o processo independentemente de como suas fases são definidas com precisão. Tais problemas são:

[...] a forma como a tarefa de projeto é organizada e dividida entre diferentes equipes de projeto ou membros da equipe; a forma como a tomada de decisões em relação ao projeto é estruturada; e a inclusão (ou exclusão) das partes interessadas no processo de projeto e a forma como elas são incluídas (ou excluídas) (Devon; Van de Poel, 2004, p. 465, tradução nossa).

O designer carrega, portanto, responsabilidade social naquilo que produz. Conforme Rodrigues (2009) o design é uma atividade de mão dupla que:

[...] ao buscar sua matéria prima na sociedade, devolve à ela um produto reelaborado, transformando o senso comum em uma expressão de linguagem mais elaborada. Desta forma o design abre novos caminhos na densidade cultural apontando rumos futuros, desencadeando relações, modelando pensamentos, dialogando com a sociedade, construindo valores, símbolos, e novos significados. Isto é interferência e, por isso mesmo, requer do designer responsabilidade social (Rodrigues, 2009, p. 11).

Sobre a responsabilidade ética em desenvolvimentos envolvendo a IA, Dignum (2019) destaca que teorias, métodos e algoritmos são necessários para integrar valores sociais, legais e éticos em todas as fases do desenvolvimento da IA e isso inclui análise, design, construção, implementação e avaliação. Assim, para projetos que envolvam o design de soluções com IA, metodologias são necessárias para extrair os valores defendidos pelos designers e pelas partes interessadas de modo a

acomodar o pluralismo de valores e entender como projetar para eficiência, usabilidade, flexibilidade, resiliência, justiça, dignidade, felicidade, bem-estar, segurança, saúde, empatia, amizade, solidariedade e paz (Dignum, 2019, p. 6, tradução nossa).

Para alcançar soluções com este nível de responsabilidade, Dignum destaca que uma postura de responsabilidade

aplica a ética ao design de IA de diferentes maneiras conforme a Figura 1:

Figura 1: Ética aplicada ao design de IA de diferentes maneiras como postura de responsabilidade nos desenvolvimentos em IA.

Ética no Design:

Refere-se aos processos regulatórios e de engenharia que apoiam o design e a avaliação de sistemas de IA à medida que estes se integram ou substituem estruturas sociais tradicionais.

Aqui, o objetivo é garantir que as equipes de desenvolvimento de sistemas de IA estejam atentas aos potenciais consequências para indivíduos e sociedades, antecipando as consequências das escolhas de design, refletindo sobre o problema que está sendo resolvido, envolvendo todas as partes interessadas, verificando e validando o design e tomando as medidas apropriadas para garantir a aceitabilidade social, legal e ética do sistema. Isso significa que precisamos perceber que os princípios de Responsabilidade, Prestação de Contas e Transparência (RPC) estão no cerne do design de sistemas de IA.

Ética por Design:

Trata da ética do comportamento dos sistemas de IA.

O trabalho nesta área se concentra em (a) os requisitos para a representação e o uso de valores morais por sistemas de IA, (b) a compreensão do significado e a especificação de restrições adequadas ao comportamento do sistema e (c) a integração de capacidades de raciocínio ético como parte dos algoritmos que determinar o comportamento de sistemas autônomos artificiais.

Ética para Designers:

Refere-se aos códigos de conduta, aos requisitos regulatórios e aos padrões e processos de certificação que garantem a integridade de todos os atores enquanto pesquisam, projetam, constroem, empregam e gerenciam sistemas de inteligência artificial.

Isso visa garantir que eles considerem o impacto social de suas escolhas de design e tomem as medidas necessárias para minimizar o impacto negativo e o uso duplo de seus resultados. Isso inclui a adesão a códigos de conduta específicos e a definição e uso de padrões, regulamentos e processos de certificação que garantem a integridade de desenvolvedores, produtos e serviços.

Fonte: Adaptado de *Dignum*, 2019, p. 6-7, tradução nossa.

Tem-se então, a partir destas três dimensões, que ética e design constituem mutuamente a base para desenvolvimentos éticos que envolvam a IA perpassando respectivamente: processos, projeto e prática profissional. É sobre esta terceira dimensão que este artigo se dedica a fazer proposições.

O CÓDIGO DE ÉTICA DO DESIGNER SEGUNDO A *WORLD DESIGN ORGANIZATION*

A *World Design Organization* (WDO) foi oficialmente fundada em 29 de junho de 1957, em uma reunião especial em Londres, sob o nome de *International Council of Societies of Industrial Design* (ICSID). A organização foi oficialmente registrada em Paris, onde a Secretaria foi estabelecida no número 17 do Quai Voltaire. As décadas de 1990 e 2000 testemunharam a expansão das ações do ICSID, que promoveu premiações, congressos e eventos inclusive em conjunto com a UNESCO. A partir da década de 2010 o ICSID desenvolveu modelos inovadores de colaboração internacional com organizações intimamente ligadas à profissão de design industrial.

Em 1º de janeiro de 2017, o ICSID tornou-se oficialmente a *World Design Organization* (WDO). Hoje a WDO é uma organização não governamental reconhecida globalmente e cujo objetivo é promover e desenvolver a disciplina do design industrial e seu poder de aprimorar a qualidade de vida econômica, social, cultural e ambiental. A WDO disponibiliza em seu site um documento oficial (com registro nas propriedades do documento da data de 2017) que traz o código de ética do profissional designer, o qual é apresentado da seguinte maneira:

O código a seguir fornece um esboço de diretrizes éticas concebidas para promover a qualidade da profissão de design industrial. Os artigos especificados neste código não devem ser considerados excludentes entre si, mas sim ser aplicados de forma consistente e holística na prática diária do design industrial (WDO, 2025).

O código traz cinco artigos que estão apresentados traduzidos na Figura 2:

Figura 2: Artigos que constituem o Código de ética dos designers da WDO.

ARTIGO 1: Beneficiar o Cliente

A responsabilidade primordial dos designers industriais para com seus clientes será concretizada através do fornecimento de projetos apropriados e originais que representem valor e benefício para seus clientes, para os clientes dos clientes e para o público em geral, ao mesmo tempo em que atendem aos objetivos éticos e empresariais dos clientes.

a) Reconhecimento dos objetivos do cliente. Os designers devem trabalhar dentro dos limites de sua profissão para promover os interesses de curto e longo prazo de seus clientes, fornecendo o seguinte:

- * Consideração pelos objetivos estratégicos, econômicos e técnicos.
- * Projetos apropriados, de alta qualidade e competitivos.
- * Melhores práticas profissionais.
- * Meios de produção eficientes, econômicos e ambientalmente corretos.
- * Práticas de negócios honestas.

b) Clareza dos acordos contratuais. Os designers definirão claramente a base sobre a qual sua remuneração total é calculada, antes de aceitarem uma atribuição.

c) Respeito pela confidencialidade do cliente. Os designers manterão absoluta confidencialidade em todos os assuntos relacionados à tecnologia, estratégia, organização e práticas de negócios de um cliente e/ou qualquer outro assunto que seja definido como confidencial pelo cliente, a menos que sejam expressamente autorizados a divulgar tais informações pelo cliente, ou se e quando tais informações se tornarem parte do domínio público.

d) Reconhecimento de conflitos pessoais/profissionais. Os designers não assumirão ou aceitarão conscientemente uma posição em que seus interesses pessoais entrem em conflito com seu dever profissional. Caso um designer se encontre nessa situação, ele deverá informar o cliente sobre esses conflitos.

e) Aumento da eficiência através da aplicação eficaz da tecnologia digital na prática de design. O designer deve aplicar a tecnologia e os recursos de Computer-Aided Industrial Design (CAID) para melhorar a qualidade do design e reduzir os riscos associados ao desenvolvimento de produtos. A implementação e o uso de um ambiente eficaz baseado na web também permitem que os designers mantenham linhas de comunicação fluidas e transmissão instantânea de dados.

ARTIGO 2: Beneficiar o Usuário

Os designers reconhecem suas contribuições para o bem-estar social, individual e material do público em geral, particularmente no que diz respeito à saúde e segurança, e não agirão conscientemente de maneira prejudicial ou contraditória a esse bem-estar.

Os designers industriais devem defender e considerar cuidadosamente as necessidades de todos os potenciais usuários, incluindo aqueles com diferentes capacidades, como os idosos e os portadores de deficiência física. Neste aspecto, os designers pensarão em toda a cadeia de valor, desde a produção até a venda e o uso do produto. Os designers percebem que a humanização da tecnologia, a ideia, a usabilidade e até mesmo o prazer de uso do produto fazem parte de sua responsabilidade.

ARTIGO 3: Proteger o Ecossistema da Terra

Em última análise, os melhores interesses das gerações atuais e futuras só podem ser protegidos se o ecossistema mundial puder ser salvaguardado. Consequentemente, os designers devem adotar os seguintes princípios de gestão ambiental:

- a) Defesa de produtos e serviços seguros. Os designers defenderão junto aos seus clientes o desenvolvimento de ambientes, paisagens, produtos, comunicações e embalagens que minimizem os danos ambientais e sejam seguros para uso por todas as pessoas.
- b) Proteção da biosfera. Os designers procurarão minimizar a liberação de qualquer poluente que possa colocar em perigo a vida, o ar, a água ou a terra.
- c) Uso sustentável de recursos naturais. Os designers se esforçarão para especificar processos e materiais que sejam o resultado de recursos naturais sustentáveis e/ou renováveis, incluindo a proteção da vegetação, habitat da vida selvagem, espaços abertos e áreas selvagens. Os designers compartilharão informações que ajudarão seus pares a fazerem as melhores escolhas ao especificar materiais e processos.
- d) Redução de resíduos e aumento da reciclagem. Os designers tentarão minimizar o desperdício. Para este fim, eles farão design visando a durabilidade, adaptabilidade, reparo e reciclagem do produto.
- e) Uso consciente de energia. Os designers escolherão fontes de energia ambientalmente seguras e adotarão meios de produção e operação que conservem energia sempre que possível.
- f) Uso de novas tecnologias. Os designers permanecerão continuamente conscientes das possibilidades oferecidas pelas novas tecnologias e as utilizarão para criar materiais que economizem recursos.

ARTIGO 4: Enriquecer a Identidade Cultural

Os designers industriais reconhecem que os ambientes, objetos e serviços criados como resultado do processo de design tanto refletem quanto ajudam a definir a identidade cultural de suas nações e das sociedades distintas dentro das nações.

Os designers devem se esforçar para incorporar e promover as tradições culturais de suas sociedades nacionais, ao mesmo tempo em que incorporam as melhores características dos princípios e padrões internacionais de design.

ARTIGO 5: Beneficiar a Profissão

Os designers devem cumprir as seguintes diretrizes para promover o desenvolvimento contínuo e o respeito pela própria profissão de design industrial:

- a) Defesa de práticas de negócios profissionais. As práticas de negócios profissionais serão fortalecidas pela aplicação dos seguintes princípios:
- O avanço profissional será baseado principalmente na qualidade do trabalho de design e não à custa de outros designers.
 - Nenhum trabalho será realizado a convite de um cliente sem a remuneração apropriada, a menos que seja para uma organização de caridade ou sem fins lucrativos.
 - O cliente será notificado com antecedência quando um designer puder se beneficiar financeiramente (por meio de associação com qualquer empresa, firma ou negócio) de quaisquer recomendações feitas por ele/ela durante a realização de um projeto, e
 - Quando solicitado a aconselhar sobre a recomendação de outros designers, um designer não aceitará pagamento de nenhuma forma dos designers selecionados.
- b) Apoio ao desenvolvimento contínuo da profissão. O desenvolvimento profissional será apoiado através do fornecimento de:
- Oportunidades para aplicar habilidades de design adquiridas em projetos desafiadores.
 - Reconhecimento justo pelo trabalho realizado.
 - Oportunidades para educação continuada.
 - Práticas de trabalho e remuneração respeitáveis para colegas, subcontratados, fornecedores etc., com os quais o trabalho foi realizado.
 - Apoio e (quando possível) participação em atividades e programas que visam aprimorar a profissão de design, seja por organizações profissionais, instituições de ensino ou órgãos legislativos.
- c) Coerência da imagem profissional. Os esforços promocionais devem refletir consistentemente os objetivos profissionais e incluir apenas informações factuais, verdadeiras, declarações e, onde apropriado, críticas objetivas e construtivas.

Fonte: Adaptado de WDO, 2017, tradução nossa.

Embora sucinto, o código trata de dimensões importantes que constituem a realidade profissional: Clientes (pessoas ou empresas que contratam o profissional); Usuário (pessoas que utilizam a solução projetada); Planeta (local que é di-

retamente impactados em todo o ciclo de vida de um produto, sistema ou serviço); Cultura (que recebe essa cultura material projetada e a reflete na sociedade); e o Design em si (domínio que projeta e por sua responsabilidade social deve fazer sua autocrítica constantemente). Considerando estas dimensões, é possível traçar um paralelo com as recomendações sobre a ética da inteligência artificial trabalhadas pela UNESCO.

AS RECOMENDAÇÕES SOBRE A ÉTICA DA INTELIGÊNCIA ARTIFICIAL

A Organização das Nações Unidas para a Educação, a Ciência e a Cultura (UNESCO) tem desempenhado um importante papel nas discussões sobre ética e IA em nível Global. Na 41ª sessão da Conferência Geral da Organização das Nações Unidas para a Educação, a Ciência e a Cultura reunida em Paris de 9 a 24 de novembro de 2021, a entidade apresentou um documento intitulado “Recomendações sobre a ética da Inteligência Artificial”.

Este documento de 44 páginas publicado em 2022 trata das diferentes dimensões nas quais o uso da IA pode levantar questões éticas. A partir destas dimensões, a publicação apresenta valores e princípios orientadores a serem seguidos pelos Estados no desenvolvimento de estruturas ou mecanismos políticos considerando os desenvolvimentos em IA e seus impactos sociais, para que estes sejam seguidos por outras partes interessadas, como, por exemplo, empresas do setor privado, instituições acadêmicas e de pesquisa e sociedade civil.

Este artigo levará em consideração para comparação com o código de ética da WDO apenas os dez princípios propostos pela UNESCO, uma vez que os valores são basilares, ou seja, devem ser contemplados por meio da operacionalização de cada um dos princípios, que no caso do design podem ser convertidos em processos de projeto ou mesmo funções nos produtos, sistemas e serviços.

De maneira resumida os valores são: Respeito, proteção e promoção dos direitos humanos, das liberdades fundamentais e da dignidade humana; Prosperidade ambiental e ecossistêmica; Garantir diversidade e inclusão e Viver em socieda-

des pacíficas, justas e interconectadas. Quanto aos princípios, estes estão apresentados de maneira sintética na Figura 3:

Figura 3: Princípios para a ética da Inteligência Artificial da UNESCO

Princípio 1: Proporcionalidade e não causar dano

Diz respeito ao cuidado para que nenhum dos processos relacionados ao ciclo de vida dos sistemas de IA exceda o necessário para atingir objetivos ou metas legítimas sendo que estes devem ser adequados a cada contexto.

Também devem ser asseguradas a implementação de procedimentos para avaliação de risco e a adoção de medidas para impedir a ocorrência de danos tanto aos seres humanos, quanto ao meio ambiente.

Princípio 2: Segurança e proteção

Destaca que em se tratando de sistemas de IA deve-se garantir proteção e segurança sendo que os danos indesejados (riscos de proteção) e as vulnerabilidades a ataques (riscos de segurança) devem ser evitados e abordados, impedidos e eliminados durante o ciclo de vida desses sistemas.

Também o desenvolvimento de marcos sustentáveis de proteção à privacidade dos dados de acesso são ações que permitirão que melhores modelos de IA sejam desenvolvidos favorecendo uma IA segura e protegida.

Princípio 3: Justiça e não discriminação

Refere-se ao cuidado que se deve ter na promoção da justiça social, equidade e não discriminação de qualquer tipo, de modo que esteja em conformidade com o direito internacional e a partir de uma abordagem inclusiva para garantir que os benefícios das tecnologias de IA estejam disponíveis e acessíveis a todos.

Obviamente estas ações devem levar em consideração as especificidades dos diferentes grupos que possam ser desfavorecidos ou marginalizados e além de atuar contra a discriminação e a determinação algorítmica tendenciosa.

Princípio 4: Sustentabilidade

Destaca que o desenvolvimento de sociedades mais sustentáveis depende também de ações no âmbito tecnológico, assim, a avaliação contínua dos impactos das tecnologias de IA nos aspectos humanos, sociais, culturais, econômicos e ambientais deve ser realizada com pleno conhecimento das implicações que tais tecnologias trazem para a sustentabilidade sendo que estes devem ser um conjunto de objetivos a serem monitorados entendendo-os como aspectos em constante evolução.

Princípio 5: Direito à privacidade e proteção de dados

Enfatiza que a privacidade é um direito essencial para proteger a dignidade, a autonomia e a capacidade de ação humanas e deve ser respeitada, protegida e promovida ao longo do ciclo de vida dos sistemas de IA.

Nesse contexto, os dados devem ser coletados, utilizados, compartilhados, arquivados e apagados em consonância com o direito internacional estando alinhados com os valores e princípios estabelecidos na Recomendação da UNESCO, respeitando também os marcos jurídicos nacionais, regionais e internacionais relevantes.

Princípio 6: Supervisão humana e determinação

Visa garantir que sempre seja possível atribuir responsabilidade ética e legal em qualquer estágio do ciclo de vida dos sistemas de IA entendendo a supervisão como não somente a supervisão humana individual, mas também a supervisão pública inclusiva desses sistemas e seus impactos.

Neste princípio entende-se que as pessoas podem recorrer à IA por motivos de eficácia para alguma necessidade ou tarefa, mas a decisão de ceder o controle em contextos limitados continua sendo de responsabilidade humana sendo, portanto, também sua a prestação de contas. Este princípio ainda destaca que decisões de vida e morte não devem ser transferidas a sistemas de IA.

Princípio 7: Transparência e explicabilidade

Destaca a importância das soluções em IA serem transparentes para os envolvidos

e explicáveis, sendo estas essenciais para garantir o respeito, a proteção e a promoção dos direitos humanos, das liberdades fundamentais e dos princípios éticos.

Na transparência se permite que as pessoas entendam o impacto e como cada estágio de um sistema é colocado em prática, considerando contexto e a sensibilidade do sistema de IA, enquanto a explicabilidade significa tornar inteligível e fornecer informações sobre o resultado dos sistemas de IA estando também relacionada com a transparência, pois os resultados devem ser compreensíveis e rastreáveis, conforme o contexto.

Princípio 8: Responsabilidade e prestação de contas

Segundo este princípio os atores de IA e os Estados-membros devem assumir responsabilidades éticas e jurídicas de acordo com o direito nacional e internacional.

Devem ser desenvolvidos mecanismos para supervisão, avaliação de impacto, auditoria e diligência, incluindo proteção de denunciantes, para garantir que exista a responsabilização pelos sistemas de IA, bem como seu impacto na sociedade.

Princípio 9: Conscientização e alfabetização

Destaca a importância e a necessidade de se promover a compreensão pública das tecnologias de IA e do valor dos dados por meio de educação aberta e acessível, engajamento cívico, habilidades digitais e treinamento em ética da IA, bem como alfabetização midiática e informacional, além de treinamento conduzido em conjunto por diferentes atores como governos, organizações intergovernamentais, sociedade civil, universidades, meios de comunicação, líderes comunitários e setor privado, sendo que a compreensão dos sistemas de IA devem ser baseadas em seu impacto nos direitos humanos, no meio ambiente e nos ecossistemas.

Princípio 10: Governança e colaboração adaptáveis e com múltiplas partes interessadas

Identifica que o direito internacional e a soberania nacional devem ser respeitados no uso de dados e que os Estados podem regulamentar os dados gerados dentro de seus territórios ou que dados que os atravessam, bem como tomar medidas para uma regulamentação de dados eficaz com base no respeito ao direito à privacidade e em conformidade com as leis internacionais e outras normas e padrões de direitos humanos.

Fonte: Adaptado de UNESCO, 2022, tradução nossa.

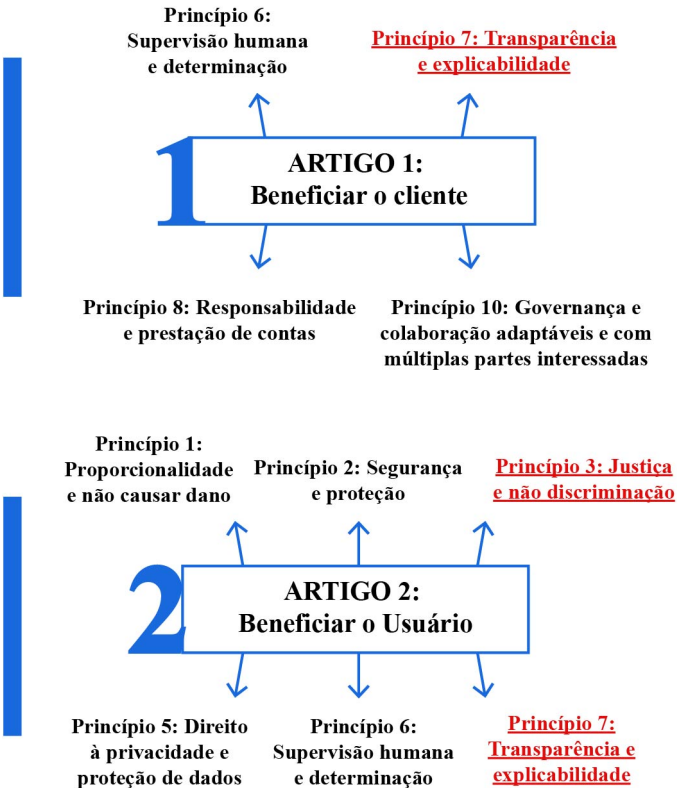
Estes princípios, segundo a publicação, devem ser promovidos, quando necessário, por meio de alterações à legislação existente e da elaboração de novas leis, regulamentos e diretrizes empresariais. Assim, entende-se que os mesmos podem pautar alterações em documentações como o código de ética dos profissionais envolvidos no desenvolvimento da IA, incluindo os designers.

O CÓDIGO DE ÉTICA DO DESIGNER REVISITADO: REFLEXÕES E PROPOSIÇÕES PARA UMA PRÁXIS ÉTICA NA ERA DA IA

Para identificar lacunas no código de ética do designer diante da realidade contemporânea na qual a IA escala, se faz necessário traçar um paralelo entre os conteúdos anteriormente apresentados. A partir da análise documental empreendida, foi possível realizar a análise de conteúdo, considerando o critério semântico (Bardin, 2011), para identificar aproximações que refletem o que o código de ética atual atende em relação aos princípios da UNESCO e as dissonâncias, ou seja, aqueles princípios que ainda falta atender dada a diferença temporal entre a publicação do código (2017) e o perfil de funcionamento da IA, escalada em nível comercial a partir de 2022 com a popularização do ChatGPT junto ao público em geral.

Importa destacar que todos os dez princípios têm relevância para a práxis do design e todos devem igualmente fazer parte das dinâmicas que se estabelecem quando algum produto, sistema ou serviço passa a ser considerado para produção, contendo ou não a IA. Contudo, quando integrada em um projeto, a IA traz elementos próprios de seu funcionamento que se constituem em novos parâmetros que devem ser considerados pelo campo do design. A Figura 4 traz visualmente o paralelo realizado a partir de cada artigo encontrado no código de ética do designer da WDO e considerando cada um dos dez princípios apresentados pela UNESCO. Alguns destes princípios se repetem porque se aproximam de mais de um dos artigos do código de ética do designer; os princípios representados na cor vermelha e sublinhados são aqueles que se enquadram em determinado artigo, mas quando se trata de considerar a IA no projeto, ainda não são contemplados de maneira adequada pelo código de ética atual:

Figura 4: Relações estabelecidas entre o Código de ética do designer da WDO e os princípios propostos pela UNESCO.



3 ARTIGO 3: Proteger o Ecosistema da Terra

Princípio 4: Sustentabilidade

4 ARTIGO 4: Enriquecer a Identidade Cultural

Princípio 3: Justiça e não discriminação

Princípio 9: Conscientização e alfabetização

Princípio 7: Transparência e explicabilidade

Princípio 8: Responsabilidade e prestação de contas

5 ARTIGO 5: Beneficiar a Profissão

Princípio 9: Conscientização e alfabetização

Princípio 10: Governança e colaboração adaptáveis e com múltiplas partes interessadas

Fonte: A autora, 2025.

No *Artigo 1: Beneficiar o Cliente*, que se refere à relação do designer com os clientes (sejam clientes esporádicos ou mesmo empresas contratantes) foram aproximados os seguintes princípios:

- Princípio 6: Supervisão humana e determinação | Algo que já ocorre na profissão, uma vez que o designer supervisiona e determina suas ferramentas. No cenário com a IA, manter esse princípio permite que o profissio-

nal tenha autonomia para determinar também os caminhos da IA que precisa utilizar quando contratado (em empresa ou na modalidade *freelance*);

- Princípio 7: Transparência e explicabilidade | Embora a transparência e explicabilidade já sejam buscados em projetos atuais, os designers ainda se deparam com decisões de negócios que por vezes não são transparentes e são responsáveis por implementar os chamados padrões enganosos do design (quando a interface é desenhada para favorecer propaganda ou a conversão de uma compra sem que o usuário tenha escolhido esse caminho, por exemplo). Em se tratando da IA, este princípio será essencial para os clientes dos designers, que precisarão prestar contas sobre como a IA opera em seus produtos, sistemas ou serviços (algo que desafia os próprios desenvolvedores da IA pela natureza opaca em que a tecnologia muitas vezes opera, com vultoso volume de operações lógicas e não deixando explícito o caminho pelo qual a IA chegou naquela resposta, um fenômeno conhecido como *black box* da IA);
- Princípio 8: Responsabilidade e prestação de contas | Essencial para uma relação salutar entre designer e cliente em projetos atuais. Saber de quem é a responsabilidade e quem deve prestar contas a quem é a base dos negócios, principalmente naqueles que envolverem a IA. Quando a IA passa a fazer parte da solução projetada será fundamental, a partir das legislações e regulamentações, deixar muito claro quem se responsabiliza e quem presta contas sobre estes desenvolvimentos e seus impactos sociais. Cabe aqui a questão: no caso de uma consequência negativa, quem será responsabilizado se algo sair do previsto?
- Princípio 10: Governança e colaboração adaptáveis e com múltiplas partes interessadas | Aspectos já muito importantes para a atuação interdisciplinar e global do design, na qual o intercâmbio de informações e produções criativas entre times já é uma realidade. Quando há a integração da IA devido às implicações jurídicas que podem sur-

gir a partir do emprego desta tecnologia, se torna essencial uma governança não somente referente a arquivos que já são uma rotina hoje, tais como arquivos projetuais e termos de confidencialidade, mas também, e principalmente, referente aos dados das pessoas que estejam envolvidos na criação e no posterior uso da solução.

Quanto ao *Artigo 2: Beneficiar o Usuário*, este foi o que apresentou um maior número de princípios aproximados, seja pelo modo atual do fazer design ou pelo que pode vir a ser com a presença cada vez maior da IA nos projetos:

- Princípio 1: Proporcionalidade e não causar dano | Tomar decisões de projeto que sejam proporcionais ao que é buscado como solução e evitar qualquer dano ao usuário, este princípio deve tomar novos contornos com a IA já que muitos dos danos podem ser não previstos;
- Princípio 2: Segurança e proteção | Também uma preocupação recorrente na carreira, trata da busca por garantir ao máximo a segurança e a proteção do usuário no contato com a solução. Em projetos com IA este cuidado deve ser redobrado por conta do uso massivo de dados que a tecnologia precisa para operar e da imprevisibilidade dos resultados;
- Princípio 3: Justiça e não discriminação | Embora designers atuem visando uma produção justa, respeitando a diversidade e não discriminando, quando se trata da IA, lidar com vieses nos dados passa a ser um novo desafio que geralmente não depende dos designers, podendo se formar na base de dados em que a IA foi treinada ou do modo como foi programada, contudo, é também responsabilidade dos designers zelar pela salvaguarda das pessoas usuárias mesmo diante desta incerteza, visando colaborar também para a transparência e explicabilidade;
- Princípio 5: Direito à privacidade e proteção de dados | Um cuidado que permeia o âmbito jurídico e de governança nas empresas, mas é também uma preocupação cotidiana do designer, uma vez que dados se tornaram

parte importante da vida das pessoas inseridas em um cenário dominado pelas tecnologias digitais operando em rede. Quando se integra a IA, este direito fundamental deve ser uma das prioridades;

- Princípio 6: Supervisão humana e determinação | A supervisão humana e determinação se refere conceder ao usuário o controle total da solução no uso, algo que já ocorre nos projetos na atualidade, mas que com a inserção da IA se torna primordial. É essencial que a decisão final na interação seja do usuário;
- Princípio 7: Transparência e explicabilidade | Embora a transparência e explicabilidade sejam rotina ao pensar no usuário em projetos no geral, quando se trata da IA isso significa garantir aos usuários uma compreensão do impacto da IA quando se utiliza a solução projetada, tornando inteligível como a tecnologia funciona para entregar aquele desempenho.

Sobre o *Artigo 3: Proteger o Ecossistema da Terra*, o princípio mais aproximado foi:

- Princípio 4: Sustentabilidade | Este princípio vem sendo um dos principais desafios para o campo do design como resultado insustentável do avanço da sociedade de consumo para a qual o design produziu. Hoje, a área dedica um campo inteiro de pesquisa (design sustentável ou design para a sustentabilidade) visando mitigar as consequências, com foco na circularidade. Na era da IA, uma nova preocupação emerge a partir do consumo energético, de recursos hídricos e da pegada de carbono que os data centers deixam para operacionalizar a IA e seu funcionamento. Coloca-se um desafio ao design que não é novo, mas que ganha novos contornos e, portanto, necessitará de novas abordagens também a partir do design para ser enfrentado.

O *Artigo 4: Enriquecer a Identidade Cultural* aproxima-se dos seguintes princípios:

- Princípio 3: Justiça e não discriminação | Enriquecer a identidade cultural é também parte do que o design ma-

terializa como solução, uma vez que cada produção do design também reflete seu tempo, suas origens e traços culturais. Entretanto, a partir do emprego da IA esta nuance cultural do design pode se perder ou simplesmente ser ignorada, devido principalmente ao modo como a IA opera: a partir de dados captados globalmente, mas que em sua maioria não refletem as especificidades de regiões que ainda permanecem marginalizadas mesmo no ambiente digital, o que termina por reduzir a pluralidade de dados e, portanto, a representatividade de pessoas, saberes, visões de mundo e culturas;

- Princípio 9: Conscientização e alfabetização | Trata principalmente de favorecer a compreensão pública da IA. Oportunizar a compreensão sobre como a tecnologia funciona significa ajudar a fazer escolhas melhores na interação com ela, o que inclui os aspectos culturais. Tal acesso garante autonomia a designers, empregadores e usuários ao identificar retornos errados, vieses e respostas falsas em relação aos aspectos culturais. Esta autonomia favorece uma interação crítica com a tecnologia emancipando estes atores via escolha consciente para tomada de decisão.

Por fim, ao *Artigo 5: Beneficiar a Profissão* foram aproximados os seguintes princípios:

- Princípio 7: Transparência e explicabilidade | Para benefício da profissão é essencial que os designers atuem com transparência e de maneira que suas decisões de projeto, bem como as ações da solução projetada sejam explicáveis aos contratantes e à sociedade. Importa destacar, contudo, que em se tratando da IA é primordial que designers consigam compreender como a IA opera para que possam projetar melhor, inclusive para considerar as consequências dos projetos contendo a IA quando acessados pela sociedade;
- Princípio 8: Responsabilidade e prestação de contas | Para beneficiar a profissão é preciso responsabilizar corretamente e definir os atores que fazem a prestação de contas. Atualmente, estes fatores são tratados em contrato, mas especificamente no caso de projetos envolvendo a

IA ter um marco regulatório e jurídico é importante para saber a quem recorrer e a quem responsabilizar pelos riscos atrelados a IA em projetos de design (uma vez que os benefícios já são buscados quando se opta por um projeto deste tipo). Tal princípio pode proteger designers e usuários das consequências de decisões políticas das empresas, que podem se originar nas dinâmicas de poder que se estabelecem em torno da tecnologia visando o lucro e sobre as quais designers não têm nenhum controle;

- **Princípio 9: Conscientização e alfabetização |** Embora designers constantemente busquem se conscientizar e alfabetizar nas tecnologias que se sucedem e impactam a profissão, no caso da IA a conscientização e alfabetização de designers é tão essencial quanto a do público em geral. Compreender os riscos e potencialidades da IA permite que os designers se apropriem do que é possível ou não com esta tecnologia e o que é recomendável ou não em sua adoção no projeto, com consciência dos riscos quando ponderados em relação aos benefícios que a IA pode trazer para a solução projetada. Este princípio também permite que designers desenvolvam uma visão crítica da tecnologia para que possam tomar decisões de maneira mais prudente;
- **Princípio 10: Governança e colaboração adaptáveis e com múltiplas partes interessadas |** A governança e a colaboração já ocorrem atualmente entre designers contratados e equipes de projeto, mas no caso de projetos com a IA, para benefício da profissão, é preciso que sejam muito claras e bem definidas as responsabilidades e a prestação de contas, uma vez que, como dito no Artigo 1, quando há a integração da IA devido às implicações jurídicas que podem surgir a partir do emprego desta tecnologia, se torna essencial uma governança não somente referente a arquivos que já são uma rotina hoje tais como arquivos projetuais e termos de confidencialidade, mas também e principalmente, referente aos dados das pessoas que estejam envolvidos na criação e no posterior uso da solução. Para benefício da profissão é essencial que a colaboração aconteça, mas esta deve ser sempre norteadada pela governança.

Conclui-se, assim, que os artigos devem ser atualizados para comportar as demandas atuais originadas pelo avanço da IA e suas consequências.

CONSIDERAÇÕES FINAIS

A área do design materializa soluções em projetos de produtos, sistemas e serviços, sendo impactada pelos âmbitos social, cultural, tecnológico, econômico e político. Como agentes de mudança na sociedade por meio do design, os designers correm o risco de prejudicá-la involuntariamente se as devidas responsabilidades não forem assumidas. Assim, projetar é práxis que deve ser norteadada por uma relação dialética e crítica entre teoria e prática, visando uma ação no mundo visando o benefício da sociedade.

A escalada da IA, verificada nos últimos anos, coloca para o campo do design o desafio de compreender como produzir eticamente dentro de um contexto que inclui um avanço tecnológico sem precedentes, o qual tem alterado o curso e a velocidade da interação entre sociedade e tecnologia (que ocorre via projeto) na medida em que a integração da IA traz benefícios, mas também novos riscos e estabelece novas dinâmicas de poder com contornos geopolíticos.

Nesse sentido, entende-se que projetar integrando a IA envolverá trabalhar diretamente com impactos sociais muitas vezes imprevisíveis. Para tal, é essencial uma práxis que tem no código de ética um guia de atuação responsável. O código de ética da WDO aqui analisado data de 2017, encontra-se, portanto, desatualizado diante dos avanços tecnológicos mais recentes, contudo, se apresenta em cinco artigos que já abarcam muitos dos pontos éticos a serem considerados no dia a dia dos designers. A partir dos dez princípios que devem ser respeitados pelos atores envolvidos no desenvolvimento de soluções contendo a IA presentes no documento de recomendações da UNESCO e aqui analisados, foi possível identificar no código de ética quais pontos já são atendidos pelo mesmo e quais ainda necessitam de alguma atualização no código por não serem contemplados de maneira adequada quando se trata da integração da IA no projeto.

A análise mostrou que o código de ética do designer já é competente para orientar a práxis, mas demanda atualização urgente para abarcar particularidades que só são possíveis de identificação quando se conhece as especificidades da tecnologia e seus impactos (e no caso da IA não são poucos). Conclui-se que, assim como a IA avança e constantemente demanda adaptações na práxis, assim deve ser também o código de ética dos designers: um documento dinâmico e atualizado constantemente, para que sirva de referência ética e proteção para designers em suas atividades diárias, uma vez que estes podem estar suscetíveis às dinâmicas de poder já mencionadas. O foco deve estar na união entre design e IA, que materializada via projeto pelas mãos dos designers, ocorra de maneira ética, promovendo uma interação crítica e emancipatória com a tecnologia, bem como a salvaguarda de direitos fundamentais da sociedade.

REFERÊNCIAS

- BARDIN, L. *Análise de conteúdo*. São Paulo: Edições 70, 2011.
- BONSIEPE, Gui. *The Disobedience of Design*. Grã-Bretanha: Bloomsbury, 2022.
- CHAN, J. Design ethics: Reflecting on the ethical dimensions of technology, sustainability, and responsibility in the Anthropocene. *Design Studies*. 54(C), 184-200. 2018. Disponível em: <https://doi.org/10.1016/j.destud.2017.09.005>. Acesso em: 13 nov. 2025.
- COECKELBERGH, M. *AI Ethics*. Cambridge, MA. MIT Press Essential Knowledge series. The MIT Press, 2020.
- COECKELBERGH, M. *Why AI Undermines Democracy and What to Do About It*. Cidade: Cambridge, UK. Polity Press, 2024.
- COULDRY, N.; MEJIAS, U. A. *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. Stanford: Stanford University Press, 2019.
- CRAWFORD, K. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven, CT: Yale University Press, 2021.
- DENIS, C. Rafael. *Design para um mundo complexo*. São Paulo: Cosac Naify, 2013.
- DEVON, R.; VAN DE POEL, I. Design Ethics: The Social Ethics Paradigm. *International Journal of Engineering Education*, 20(3), 461-469, 2004. Disponível em:

https://www.semanticscholar.org/paper/Design-Ethics%3A-The-Social-Ethics-Paradigm*-Devon-Poel/17808e45db3d7d0d10f42155cb7c810b99021e77. Acesso em: 18 nov. 2025.

DIGNUM, V. *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*. Springer, 2019.

FLORIDI, L. *A ética da inteligência artificial: princípios, desafios e oportunidades*. Curitiba: PUCPRESS, 2024.

FREIRE, P. *Conscientização, teoria e prática da libertação*. 3ª ed. São Paulo, Ed.Moraes, 1980.

FRY, T. The voice of sustainment: Design ethics as futuring. *Design Philosophy Papers*, vol.2, no.2, pp. 145-156. 2004. Disponível em: <https://doi.org/10.2752/144871304X13966215068038> . Acesso em: 18 nov. 2025.

FRY, T. *Defuturing: A New Design Philosophy*. Grã-Bretanha: Bloomsbury, 2020.

NOBLE, S. U. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: New York University Press, 2018.

NORMAN, D. *Design for a Better World: Meaningful, Sustainable, Humanity Centered*. London: MIT Press, 2023.

RODRIGUES, Irene G. O olhar antropológico do designer. In: RODRIGUES, Irene G. *Faces do design 2: ensaios sobre arte, cultura visual, design gráfico e as novas mídias*. São Paulo: Edições Rosari, 2009. p. 81-97.

SILVA, T. *Racismo algorítmico: inteligência artificial e discriminação nas redes digitais*. São Paulo: Edições Sesc São Paulo, 2022.

SULEYMAN, M.; BHASKAR, M. *The Coming Wave: Technology, Power, and the Twenty-First Century's Greatest Dilemma*. New York: Crown Publishing Group, 2023.

UNESCO. Organização das Nações Unidas para a Educação, a Ciência e a Cultura. *Recomendações sobre a ética da inteligência artificial*. 2022. Disponível em: https://unesdoc.Unesco.org/ark:/48223/pf0000381137_por . Acesso em: 22 nov. 2025.

WORLD DESIGN ORGANIZATION (WDO). *Code Of Professional Ethics*. 2017. Disponível em: https://wdo.org/wp-content/uploads/WDO_CodeofEthics.pdf. Acesso em: 18 nov.2025.

**RESPONSABILIDADE NA ERA DA IA:
TRANSPARÊNCIA E MERCOSUL**

**RESPONSIBILITY IN THE AI AGE:
TRANSPARENCY AND MERCOSUL**

DOI: 10.29327/5798915.1-4

Daniela América da Silva⁴
Johnny Cardoso Marques⁴
Delmo Mattos da Silva⁴

Resumo: A delegação de decisões à Inteligência Artificial (IA) introduz um dilema de responsabilidade crítico para os seres humanos. Esse desafio é exacerbado pela velocidade e complexidade da cadeia de desenvolvimento e uso (o problema das “muitas mãos” e “muitas coisas”) e, fundamentalmente, pela falta de conhecimento e explicabilidade (XAI) inerente às operações da IA. A Transparência é identificada como fator mitigador essencial. Esta pesquisa busca definir a transparência, avaliar seu papel na redução de vieses algorítmicos e analisar sua abordagem específica nos princípios emergentes de IA dos países membros do Mercosul, propondo, por fim, uma visão macro dos requisitos de transparência em toda a região.

Palavras-chave: Inteligência Artificial, Responsabilidade, Transparência, Mercosul, Vieses Algorítmicos, Princípios

Abstract: Delegating decisions to Artificial Intelligence (AI) introduces a critical responsibility dilemma for humans. This challenge is exacerbated by the speed and complexity of the development and usage chain (the “many hands” and “many things” problem), and fundamentally, by the lack of

⁴ Instituto Tecnológico de Aeronáutica, São José dos Campos, SP, Brasil, damerica@ita.br, johnny@ita.br, delmo@ita.br.

knowledge and explainability (XAI) inherent in AI operations. Transparency is identified as the essential mitigating factor. This research seeks to define transparency, evaluate its role in reducing algorithmic bias, and analyze its specific approach within the emerging AI principles of Mercosul member countries, ultimately proposing a macro vision of the transparency requirements across the region.

Keywords: Artificial Intelligence, Accountability, Transparency, Mercosul, Algorithmic Biases, Principles

INTRODUÇÃO

A utilização da Inteligência Artificial (IA) na tomada de decisões, que possibilita a delegação de tarefas à máquina em uma extensão superior à habitual, suscita um problema central: a atribuição de responsabilidade (Coeckelbergh, 2024). Para os seres humanos, a causalidade de um efeito sobre o mundo e sobre terceiros implica a responsabilidade pelas consequências, o que, segundo Aristóteles, constitui a condição primordial para a responsabilidade moral. No entanto, ao delegar decisões e ações à IA, cujas consequências podem ser eticamente significativas, depara-se com um dilema. Uma vez que a IA carece de consciência e de raciocínio moral, ela não pode ser considerada responsável pelos seus atos (Coeckelbergh, 2024).

Máquinas podem ser consideradas agentes, mas não agentes morais, visto que lhes são inerentes a falta de consciência e de livre-arbítrio. Consequentemente, os seres humanos delegam a agência, mas retêm a responsabilidade. Não obstante, a velocidade de processamento da IA e a complexidade do seu histórico de desenvolvimento dificultam o processo de atribuição de responsabilidade (Coeckelbergh, 2024). A título de exemplo, uma IA pode ter sido concebida no âmbito de um projeto científico universitário e, subsequentemente, ser implementada em um setor de saúde ou em um contexto militar, levantando a questão: quem será o responsável? (Coeckelbergh, 2024).

Os sistemas de tecnologia implicam os conceitos de “muitas mãos” e “muitas coisas”. O cenário de “muitas mãos” estabelece que a responsabilidade por falhas, a exemplo de um acidente envolvendo um veículo autônomo, é di-

fusa. Esta abrange programadores, a empresa desenvolvedora, o usuário, órgãos reguladores, entre outros agentes. Por sua vez, o conceito de “muitas coisas” refere-se à multiplicidade de componentes técnicos: algoritmos que interagem com sensores, dados, *hardware* e *software*. No contexto do aprendizado de máquina (*machine learning*), há diversas fases (coleta, tratamento e treinamento), cada uma delas com elementos técnicos e decisões humanas, o que complexifica a causalidade (Coeckelbergh, 2024).

A responsabilidade humana pelas ações da IA está intrinsecamente conectada ao problema do conhecimento: para ser considerado responsável, o ser humano necessita saber e ser capaz de explicar o funcionamento da IA. A responsabilidade exige, portanto, elucidação e explicabilidade (Coeckelbergh, 2024). A dificuldade reside no problema da transparência. Em certos sistemas, o processo de tomada de decisão da IA não é claro nem explicável, configurando o que se denomina caixa-preta. Quando o mecanismo pelo qual a IA atinge um resultado não pode ser rastreado ou justificado, torna-se um desafio responsabilizar o ser humano, uma vez que este se vê impedido de “responder pelas ações” do sistema (Coeckelbergh, 2024).

Diante do exposto, e com o objetivo de elucidar o papel da transparência e os problemas correlatos no contexto da IA, propõem-se as seguintes Questões de Pesquisa (QPs):

- QP1: O que é a transparência da IA?
- QP2: Como a transparência auxilia na mitigação de vieses?
- QP3: Como a transparência é abordada nos princípios para a IA nos países do Mercosul?

Em consonância com as questões de pesquisa estabelecidas, o trabalho será estruturado da seguinte forma: primeiramente, a seção *Contexto* abordará a compreensão do conceito de transparência e seus desafios. Em seguida, a seção *Discussão* apresentará a relevância da transparência para a mitigação de vieses e o desenvolvimento desse conceito nos países do Mercosul. Posteriormente, a seção *Modelo Proposto* delineará uma visão macro dos principais requisi-

tos para a transparência da IA aplicáveis aos países do Mercosul. Por fim, a seção *Conclusão* resumirá as principais contribuições e realizações do estudo conduzido.

CONTEXTO

Considerando o jogo de xadrez, embora os programadores detenham o conhecimento sobre o funcionamento da IA, eles desconhecem o modo como esta concebeu uma jogada específica. Tal fato representa um problema de responsabilidade, pois inviabiliza que os seres humanos expliquem uma decisão particular do sistema (Coeckelbergh, 2024). Essa dificuldade inerente a alguns algoritmos de IA é denominada problema da caixa-preta, visto que, embora os programadores originais conheçam o código, os desenvolvedores subsequentes e os usuários não têm conhecimento preciso do que a IA está, de fato, executando (Coeckelbergh, 2024).

A transparência da IA está intrinsecamente relacionada aos conceitos de explicabilidade e interpretabilidade, uma vez que estes fornecem *insights* que auxiliam na resolução do problema da caixa-preta. Tais conceitos também contribuem para a resposta da Questão de Pesquisa (QP1), que indaga: “O que é a transparência da IA?” (Jonker *et al.*, 2025). É importante notar que estes conceitos apresentam definições e casos de uso distintos:

- **Explicabilidade da IA:** Busca responder: “Como o modelo chegou a estes resultados?” (Jonker *et al.*, 2025). É o conjunto de métodos que permite aos usuários humanos compreender e confiar nas saídas geradas. A explicabilidade examina o caminho percorrido para um resultado específico, complementando a interpretabilidade ao analisar o processo real da decisão.
- **Interpretabilidade da IA:** Visa responder: “Como o modelo toma decisões?” (Jonker *et al.*, 2025). Torna todo o processo de funcionamento da IA compreensível, medindo a taxa de sucesso com que humanos podem prever a saída do sistema, focando na lógica subjacente, relevância e consequências esperadas.

- **Transparência da IA:** É o conceito mais abrangente, buscando responder: “Como o modelo foi criado, quais dados o treinaram e como ele toma decisões?” (Jonker *et al.*, 2025). Ela engloba todos os fatores do desenvolvimento e implementação, incluindo dados de treinamento e políticas de acesso.

Soma-se a isso um problema de conhecimento relacionado à IA, dado que muitos usuários desconhecem o funcionamento do sistema, seus potenciais efeitos e, em diversas ocasiões, sequer estão cientes de sua utilização (Coeckelbergh, 2024). Tais problemas podem, adicionalmente, ser avaliados sob a perspectiva da confiança, visto que a ausência de transparência culmina na diminuição da credibilidade tanto na tecnologia quanto nos indivíduos que a empregam (Coeckelbergh, 2024).

A transparência é crucial para a reflexão social, pois o risco não é apenas a dominação por uma elite tecnológica, mas uma sociedade tecnológica onde ninguém se responsabiliza pelos impactos gerados (Jonker *et al.*, 2025). A abertura da “caixa-preta” beneficia a tecnologia, permitindo a identificação e eliminação de correlações espúrias, aprimorando o sistema (Jonker *et al.*, 2025). Contudo, a implementação da transparência é desafiadora. Instituições frequentemente evitam revelar seus algoritmos para proteger interesses comerciais e propriedade intelectual, o que pode comprometer a divulgação necessária (Jonker *et al.*, 2025). O equilíbrio entre o benefício social e a proteção comercial é, portanto, central.

A explicabilidade da IA, um pilar da transparência, transcende a mera comunicação de decisões, levantando questões filosóficas e científicas sobre a natureza da explicação em si. Estudos em psicologia e ciência cognitiva (Miller, 2019) sugerem que, como nas interações humanas, as explicações da IA devem ser sociais, adaptadas às crenças do receptor, em vez de um encadeamento causal completo.

A transparência é vital para a IA Responsável, pois permite aos *stakeholders* avaliar a precisão preditiva, a imparcialidade e a mitigação de vieses nos modelos (Jonker *et al.*, 2025; da Silva *et al.*, 2024; da Silva e Marques, 2025).

Essa abordagem visa alinhar o desenvolvimento, projeto e implementação da IA com padrões legais, éticos e os interesses das partes envolvidas.

Considerando que os sistemas de IA têm uso regional ou global, a transparência exige o conhecimento dos princípios de diversas tradições culturais. Compreender essas nuances é crucial para a aplicabilidade da explicabilidade da IA em diferentes contextos, fomentando uma explicação socialmente contextualizada da tecnologia (da Silva *et al.*, 2025; Coeckelbergh, 2024).

Neste âmbito, o viés configura-se como um problema de natureza tanto ética quanto social. Quando uma IA toma ou recomenda decisões, estas podem se manifestar como injustas ou desleais para indivíduos ou grupos específicos. Embora o viés possa surgir em aplicações de IA clássica – por exemplo, em um sistema especialista ou um banco de dados –, ele está prevalentemente associado a aplicações de aprendizado de máquina (Coeckelbergh, 2024). A compreensão desse fenômeno propicia a resposta à Questão de Pesquisa (QP2): “Como a transparência auxilia na mitigação de vieses?”

O viés, via de regra, não é intencional, manifestando-se frequentemente devido à falha dos desenvolvedores da IA em prever efeitos discriminatórios contra grupos ou indivíduos específicos. Essa ocorrência se deve a múltiplos fatores, como a insuficiente compreensão do sistema de IA, a falta de ciência acerca do problema de enviesamento, o desconhecimento dos próprios preconceitos ou a incapacidade de antecipar as consequências não intencionais da tecnologia. Adicionalmente, observa-se a insuficiência de reflexão sobre tais consequências e a ausência de contato com as partes interessadas relevantes (*stakeholders*) (Coeckelbergh, 2024; da Silva *et al.*, 2024).

Entretanto, os impactos decorrentes de decisões tendenciosas da IA podem gerar consequências graves para o acesso a recursos e para as liberdades individuais. Indivíduos podem ser impedidos de obter emprego ou crédito, ser submetidos à prisão ou sofrer violência. Ademais, os efeitos de tais decisões não se restringem aos indivíduos, podendo

afetar comunidades inteiras (Coeckelbergh, 2024; da Silva *et al.*, 2021; da Silva *et al.*, 2023).

Um caso notório é o do sistema COMPAS (utilizado para prever a reincidência de réus na Flórida). Uma análise conduzida pela ProPublica demonstrou que o sistema produziu uma taxa desproporcional de falsos positivos para réus negros (aqueles previstos para reincidir, mas que não o fizeram) e uma taxa desproporcional de falsos negativos para réus brancos (aqueles previstos para não reincidir, mas que reincidiram), o que contribuiu para a perpetuação de disparidades raciais (Coeckelbergh, 2024).

Outra fonte de viés reside nas bases de dados, as quais podem carecer de representatividade da população global. A base ImageNet, a título de exemplo, apresenta uma forte representação dos Estados Unidos da América (EUA), enquanto países como China e Índia estão sub-representados. Este desequilíbrio é atribuído ao fato de as imagens terem sido coletadas por meio de buscas tendenciosas (Coeckelbergh, 2024; Yang *et al.*, 2020).

Vieses de gênero também se manifestam, sendo evidentes, por exemplo, em Modelos de Linguagem Natural (LLMs) aplicados à área da saúde. Estudos apontam que a linguagem gerada para descrever as necessidades clínicas femininas demonstrou maior propensão a minimizar as condições das pacientes, em contraste com a linguagem mais direta empregada para o público masculino. Tal discrepância pode resultar em disparidades no acesso a serviços de saúde, visto que estes são alocados com base na necessidade clínica (Coeckelbergh, 2024; Rickman, 2025).

Por fim, há o viés relacionado ao uso humano da IA, que envolve o risco de confiança excessiva nos algoritmos, levando os tomadores de decisão a desconsiderar seu próprio julgamento. Em modelos de linguagem, essa confiança exacerbada pode se manifestar como comportamentos de apego excessivo, nos quais a IA valida em demasia os sentimentos do usuário, o que pode prejudicar a saúde mental (Coeckelbergh, 2024; O'Donnell, 2025; Slattery *et al.*, 2024).

Contudo, um algoritmo totalmente imparcial é inviável, pois não há consenso sobre justiça ou equidade perfeita. Os dados da IA são abstrações e refletem escolhas humanas, não sendo neutros (Coeckelbergh, 2024). Assim, o tratamento do viés na IA é uma questão política e filosófica, ligada ao modelo de sociedade desejado. Se a decisão humana é injusta, a IA poderia ter o papel de revelar nossos preconceitos inerentes e instruir-nos sobre a condição humana (Coeckelbergh, 2024).

DISCUSSÃO

O arcabouço regulatório em torno da IA encontra-se em constante evolução, sendo a Lei de IA da União Europeia (UE) o primeiro *framework* abrangente a nível global. Esta legislação adota uma abordagem baseada em risco, que implica a aplicação de regras distintas e, em certos casos, a proibição de usos específicos. Para a devida conformidade, a lei exige transparência, governança rigorosa e uma eficaz gestão de riscos (Jonker *et al.*, 2025).

Especialistas preveem um “Efeito Bruxelas”, análogo ao Regulamento Geral de Proteção de Dados (GDPR), onde a Lei de IA da UE atuará como catalisadora para o desenvolvimento de padrões globais de ética e governança. Embora a maioria dos países ainda careça de legislação abrangente, a existência de *frameworks* e de projetos de lei em blocos regionais como o Mercosul evidencia que o tema é incipiente, mas está em ascensão. A transparência dos processos é, portanto, crucial para atender às exigências de auditores e órgãos reguladores (Jonker *et al.*, 2025). Exemplos relevantes no Mercosul são sumarizados no Quadro 1, que demonstra os projetos de lei que foram recentemente aprovados ou estão em tramitação nos países do bloco (Blanchet, 2025).

Existe, ainda, uma declaração conjunta dos países do bloco, publicada em 2023, intitulada “Mercosul– Declaração de Princípios de Direitos Humanos sobre Inteligência Artificial” (Mercosul, 2023). No que concerne à transparência, a declaração estabelece que:

A transparência e a explicabilidade algorítmica, bem como o controle humano da tecnologia, são essenciais

para garantir que as decisões tomadas pelos sistemas de inteligência artificial sejam compreensíveis para as partes envolvidas, afetadas e interessadas. Os sistemas de inteligência artificial devem ser projetados de maneira a permitir uma supervisão adequada, de acordo com as normas de cada país, e a explicação de seus processos e resultados, de forma universal por idôneos na matéria.

O documento enfatiza a importância da supervisão adequada em conformidade com a legislação interna de cada nação.

Quadro 1 - Principais projetos de lei no Mercosul e princípios considerados.

País	Situação da Lei de Inteligência Artificial
Argentina	Iniciativas legislativas em desenvolvimento, com ênfase na proteção de dados pessoais e na incorporação de princípios de transparência, equidade e proteção de dados, ainda sem marco regulatório específico para IA (Argentina, 2023).
Brasil	Marco consolidado de proteção de dados por meio da LGPD, complementado por estratégias nacionais de IA, e projeto de lei em tramitação no Congresso Nacional voltado à regulação do uso da inteligência artificial (Brasil, 2024; 2025).
Chile	Projeto de Lei de Inteligência Artificial e Robótica em tramitação, com enfoque na regulação de sistemas de IA a partir dos princípios de não discriminação, inclusão e transparência (Chile, 2025).
México	Propostas legislativas em tramitação que articulam a proteção de dados pessoais e os direitos humanos, com iniciativas setoriais voltadas à regulação da inteligência artificial (Mexico, 2024a; 2024b).
Peru	Marco normativo vigente que inclui a Estratégia Nacional de Inteligência Artificial e legislação específica, orientada por princípios de sustentabilidade, inclusão e direitos humanos (Peru,2023).
Colômbia	Legislação de proteção de dados pessoais vigente, acompanhada de propostas normativas em desenvolvimento para a construção de um regime regulatório específico de inteligência artificial (Colombia, 2023; 2025).

Equador	Projeto de Lei de Governança da Inteligência Artificial em tramitação, com foco em ética, segurança e proteção de dados no uso de sistemas de IA (Equador, 2025).
Uruguai	Estratégia nacional e iniciativas legislativas em tramitação que tratam da inteligência artificial e das neurotecnologias, fundamentadas nos princípios de IA responsável, neurodireitos e não discriminação (Uruguai, 2025).

Dada a relevância de se compreender as normativas internas de cada país, a seção subsequente apresentará a descrição da transparência nas principais legislações das nações do Mercosul. O objetivo é, assim, responder à Questão de Pesquisa (QP3): “Como a transparência é abordada nos princípios para a IA nos países do Mercosul?”

ARGENTINA

Embora não exista uma regulamentação específica para a IA, a norma 87/2023 estabelece diretrizes para a proteção de dados. O Artigo 6º define o Princípio de licitude, lealdade e transparência, nos seguintes termos:

- “Os dados pessoais devem ser tratados de maneira lícita, leal e transparente. O tratamento é considerado lícito se for realizado de acordo com o estabelecido na presente Lei e complementaridade normativa. Considera-se leal se o responsável pelo tratamento se abster de tratar os dados através de meios enganosos ou fraudulentos. É transparente se as informações que estiverem vinculadas ao tratamento dos dados forem facilmente acessíveis e utilizarem uma linguagem simples e clara” (Argentina, 2023).

BRASIL

No que se refere à Lei Geral de Proteção de Dados (LGPD) do Brasil (Brasil, 2024), o princípio da transparência é estabelecido nos seguintes artigos:

- Art. 6º As atividades de tratamento de dados pessoais deverão observar a boa-fé e os seguintes princípios: VI – Transparência: Garantia, aos titulares, de informações claras, precisas e facilmente acessíveis sobre a realiza-

ção do tratamento e os respectivos agentes de tratamento, observados os segredos comercial e industrial.

- Art. 9º O titular tem direito ao acesso facilitado às informações sobre o tratamento de seus dados, que deverão ser disponibilizadas de forma clara, adequada e ostensiva acerca de, entre outras características previstas em regulamentação para o atendimento do princípio do livre acesso:
 - § 1º Na hipótese em que o consentimento é requerido, este será considerado nulo caso as informações fornecidas ao titular tenham conteúdo enganoso ou abusivo ou não tenham sido apresentadas previamente com transparência, de forma clara e inequívoca.
- Art. 10 O legítimo interesse do controlador somente poderá fundamentar tratamento de dados pessoais para finalidades legítimas, consideradas a partir de situações concretas, que incluem, mas não se limitam a:
 - § 2º O controlador deverá adotar medidas para garantir a transparência do tratamento de dados baseado em seu legítimo interesse.
- Art. 40 A autoridade nacional poderá dispor sobre padrões de interoperabilidade para fins de portabilidade, livre acesso aos dados e segurança, assim como sobre o tempo de guarda dos registros, tendo em vista especialmente a necessidade e a transparência.

Em relação ao Projeto de Lei (PL) nº 2.338/2023 (Brasil, 2025), destacam-se os seguintes pontos sobre a transparência:

- Art. 2º O desenvolvimento, a implementação e o uso de sistemas de inteligência artificial no Brasil têm como fundamentos:
 - VI – Transparência, explicabilidade, inteligibilidade e auditabilidade.
- Art. 18. Caberá à autoridade competente atualizar a lista dos sistemas de inteligência artificial de risco excessivo ou de alto risco, identificando novas hipóteses, com base em, pelo menos, um dos seguintes critérios:

- VII – Baixo grau de transparência, explicabilidade e auditabilidade do sistema de inteligência artificial, que dificulta o seu controle ou supervisão.
- Art. 19. Os agentes de inteligência artificial estabelecerão estruturas de governança e processos internos aptos a garantir a segurança dos sistemas e o atendimento dos direitos de pessoas afetadas, nos termos previstos no Capítulo II desta Lei e da legislação pertinente, que inclui, pelo menos:
 - I – Medidas de transparência quanto ao emprego de sistemas de inteligência artificial na interação com pessoas naturais, o que inclui o uso de interfaces ser-humano-máquina adequadas e suficientemente claras e informativas;
 - II – Transparência quanto às medidas de governança adotadas no desenvolvimento e emprego do sistema de inteligência artificial pela organização.
- Art. 24. A metodologia da avaliação de impacto conterà, ao menos, as seguintes etapas:
 - i) Medidas de transparência ao público, especialmente aos potenciais usuários do sistema, a respeito dos riscos residuais, principalmente quando envolver alto grau de nocividade ou periculosidade à saúde ou segurança dos usuários, nos termos dos artigos 9 e 10 da Lei nº 8.078, de 11 de setembro de 1990 (Código de Defesa do Consumidor).

Justificativa: Além de estabelecer direitos básicos e transversais para qualquer contexto de interação entre máquina e ser humano, como informação e transparência, tal obrigação é intensificada quando o sistema de IA produz efeitos jurídicos relevantes ou impacta os sujeitos de maneira significativa (por exemplo: direito de contestação e intervenção humana).

No que tange à governança dos sistemas, o projeto de lei elenca as medidas a serem adotadas para garantir a transparência e a mitigação de vieses; estabelece medidas adicionais para sistemas de alto risco e para sistemas gover-

namentais de inteligência artificial; e normatiza o procedimento para a avaliação de impacto algorítmico.

CHILE

O Boletim n.º 15.869-19 e 16.821-19 (Chile, 2025) constitui o projeto de lei que visa regulamentar os sistemas de IA, do qual se extraem as seguintes disposições:

- **Transparência e Identificação**
 - Os sistemas de IA serão desenvolvidos e utilizados de modo a facilitar uma rastreabilidade adequada, em conformidade com o ordenamento jurídico vigente. Adicionalmente, o sistema deverá identificar-se como agente artificial em cada oportunidade de interação com seres humanos, permitindo que estes tomem conhecimento de forma clara e precisa e estejam conscientes de que a comunicação ou interação ocorre com um sistema de IA.
- **Explicabilidade**
 - Os sistemas de IA serão criados, desenvolvidos, inovados, implementados e utilizados de maneira que seus resultados ou *outputs* sejam compreensíveis e inteligíveis para as pessoas por eles impactadas; e promoverão a transparência e a rastreabilidade em todas as suas operações.
- **Art. 5º Obrigações de transparência em determinados sistemas de IA.** Todo operador de sistemas de IA que gere conteúdo sintético de áudio, imagem, vídeo ou texto deverá zelar para que seus resultados ou *outputs* sejam identificáveis como gerados ou manipulados de maneira artificial.
- **Riscos**
 - Os sistemas de IA de risco limitado devem garantir condições de transparência e segurança proporcionais ao seu nível de risco, de modo que os indivíduos sejam informados de forma clara e precisa e possam reconhecer que estão interagindo com um sistema de IA.

MÉXICO

No que diz respeito à proteção de dados (México, 2024a), o princípio da transparência é contemplado nos seguintes artigos:

- Art. 52. Para o tratamento de dados pessoais em serviços, aplicações e infraestrutura na denominada computação em nuvem (*cloud computing*), em que o responsável adere aos mesmos mediante condições ou Cláusulas Gerais de Contratação, somente será permitido utilizar estes serviços nos fornecedores que: [...] demonstrem transparência nas subcontratações que envolvam informações sobre quem presta o serviço.
- Art. 80. Os esquemas de autorregulação podem ser traduzidos em códigos deontológicos ou de boas práticas profissionais, selos de confiança, políticas de privacidade, regras de privacidade corporativas e outros mecanismos, os quais incluirão regras ou padrões específicos e atenderão aos seguintes objetivos primordiais:
 - VIII. Promover o compromisso dos responsáveis com a entrega de informações e a adoção de políticas internacionais consistentes com critérios externos, bem como auspiciar mecanismos para implementar políticas de privacidade, incluindo ferramentas, transparência, supervisão interna contínua, avaliações de risco, verificações externas e sistemas de remediação...

No que tange à Inteligência Artificial (IA) (México, 2024b), a legislação faz referência ao tema da seguinte forma:

- Em conformidade com a recomendação da UNESCO, busca-se garantir que as transformações digitais promovam os direitos humanos e contribuam para a conquista dos Objetivos de Desenvolvimento Sustentável, abordando questões relacionadas à transparência, à transmissão de informações e à privacidade. Para tal, são incluídos capítulos com políticas orientadas para a ação sobre governança de dados, educação, cultura, trabalho, atenção sanitária e economia.

- Art. 9º Todos os membros do Conselho são obrigados a observar as disposições contidas na Constituição Política dos Estados Unidos Mexicanos em matéria de serviço público, transparência, administração pública, direitos humanos, inclusão e perspectiva de gênero, bem como cumprir as demais leis que dela emanam.

PERU

Em relação à Inteligência Artificial (IA) no Peru (Peru, 2023), destacam-se os seguintes artigos:

- Art. 1º O objeto da presente Lei consiste em promover o uso da inteligência artificial no âmbito do processo nacional de transformação digital, privilegiando a pessoa e o respeito aos direitos humanos. O fim é fomentar o desenvolvimento econômico e social do país, em um ambiente seguro que garanta seu uso ético, sustentável, transparente, replicável e responsável.
- Art. 4º A Autoridade Nacional, no contexto da transformação digital, desenvolve e articula ações para promover e impulsionar:
 - (e) A adoção de diretrizes éticas para um uso sustentável, transparente e reapplicável da inteligência artificial.

COLÔMBIA

No que se refere à proteção de dados na Colômbia (Colômbia, 2023), o princípio da transparência é apresentado no seguinte artigo:

- Art. 4º Princípios para o tratamento de dados pessoais.
 - Princípio da Transparência: O Tratamento deve garantir o direito do Titular de obter, junto ao Responsável pelo Tratamento ou ao Encarregado do Tratamento, a qualquer momento e sem restrições, informações sobre a existência de dados que lhe digam respeito.

No que concerne à Inteligência Artificial (IA) na Colômbia (Colômbia, 2025), a regulamentação estabelece o seguinte:

- Art. 3º Princípios diretores para o desenvolvimento e uso da Inteligência Artificial:
 - 3.4. Ética e sustentabilidade: Promover o desenvolvimento e o uso da IA com base em princípios éticos que respeitem e protejam os direitos humanos, garantindo a transparência, a equidade, a não discriminação e a proteção da privacidade. Além disso, a ética e a solidez almejadas fomentam o equilíbrio por meio do projeto de tecnologias que minimizem seu impacto ambiental, promovam o desenvolvimento social e econômico inclusivo e priorizem o bem-estar das gerações presentes e futuras.
 - 3.5. Transparência e explicação: Os sistemas de IA devem ser projetados e utilizados com princípios que garantam a clareza, a acessibilidade e a rastreabilidade dos processos e decisões inteligentes. Tais sistemas devem garantir que suas decisões sejam compreensíveis para os usuários e autoridades reguladoras, facilitando sua verificação, auditoria e capacidade de questionar seus resultados.
- Art. 4º Definições:
 - Sistemas de alto risco: Sistemas de IA que podem afetar o exercício dos direitos à privacidade, liberdade de expressão, transparência ou acesso à informação pública, a qualquer momento de sua implementação ou posterior.
 - Sistemas de Inteligência Artificial de risco limitado: São sistemas que, sem implicar uma ameaça significativa a direitos ou segurança [...] exigem o cumprimento de obrigações de transparência e oferecem informações claras ao usuário sobre a natureza artificial do sistema e permitem sua desativação.
 - Sistemas de Inteligência Artificial de baixo risco: São sistemas com um risco mínimo para a segurança ou os direitos dos indivíduos e, por isso, estão sujeitos a uma regulamentação focada

na gestão do risco, princípios gerais de ética, transparência e boas práticas.

- Princípios e regras para a governança e o uso responsável pela Inteligência Artificial na administração pública:
 - As decisões tomadas com o apoio de sistemas de IA devem ser claras, compreensíveis e transparentes para os cidadãos, permitindo sua revisão e questionamento quando necessário. O uso destes sistemas não exime os servidores públicos nem as autoridades competentes de responderem por suas ações, nos termos estabelecidos pela lei.

EQUADOR

Em relação à Inteligência Artificial (IA) no Equador (Equador, 2025), a regulamentação apresenta as seguintes diretrizes:

- Exposição de Motivos e Princípios
 - Transparência, explicação e transmissão de informações: O funcionamento dos sistemas de IA deve ser compreensível, auditável e explicável para as pessoas afetadas. Serão promovidos mecanismos de publicação dos dados, modelos e especificações utilizados, e será exigida aos operadores de sistemas de alto risco a documentação detalhada de seus processos de projeto e funcionamento.
- Direito dos Titulares
 - Transparência e explicabilidade: Garante o direito a conhecer quando uma decisão é tomada por IA ou com a assistência de IA, a compreender seus fundamentos e a solicitar revisão humana.
- Setor Público e Acesso à Informação
 - Avaliação de impacto (sistemas de alto risco): Os responsáveis pelos sistemas de IA sujeitos a estas avaliações deverão publicar proativamente em seus sites ou canais de comunicação, em formatos abertos e acessíveis, um resumo executivo e outras informações relevantes sobre a avaliação.

- Aplicação Geral
 - Esta Lei será aplicada de forma complementar e sem prejuízo das obrigações estabelecidas na normativa nacional vigente em matéria de proteção de dados, transparência e acesso à informação pública, defesa do consumidor, competência e direito administrativo.

URUGUAI

Em relação à Inteligência Artificial (IA) no Uruguai (Uruguai, 2025), a legislação estabelece o seguinte:

- A estratégia nacional para IA deverá ser fundamentada nos princípios de equidade, não discriminação, responsabilidade, transmissão de informações, transparência, auditoria e inovação segura, com o devido respeito à dignidade humana, ao sistema democrático e à forma republicana de governo. Os princípios de proteção de dados pessoais, contidos na Lei n.º 18.331, de 11 de agosto de 2008, farão parte integrante da estratégia mencionada.
- Como componente dessa estratégia nacional de inteligência artificial, é estabelecido um prazo de cento e oitenta dias para a apresentação ao Poder Legislativo de um relatório e de recomendações para sua regulamentação legal, visando seu desenvolvimento ético, a proteção dos direitos humanos e, simultaneamente, o fomento da inovação tecnológica.

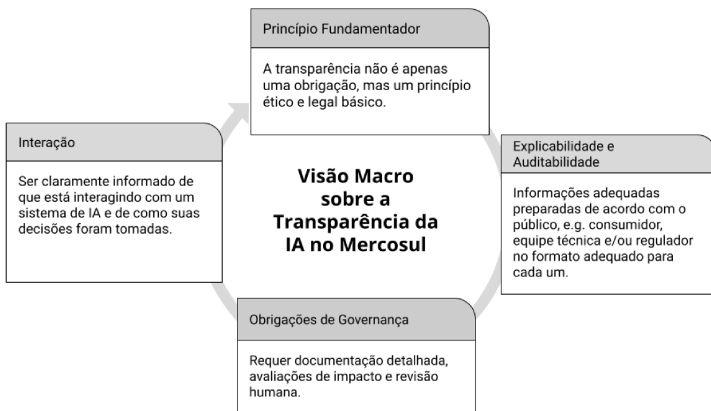
CONSIDERAÇÕES SOBRE O MERCOSUL

O denominador comum mais evidente para a governança da IA no Mercosul reside no foco na Transparência, desdobrado em quatro pilares principais, conforme detalhado a seguir e ilustrado na Figura 1, que apresenta a visão macro sobre a transparência da IA no bloco:

- Princípio Fundamentador: A transparência é estabelecida não apenas como uma obrigação, mas como um princípio ético e legal básico (Brasil, Peru, Colômbia, Equador, Uruguai).

- **Explicabilidade e Auditabilidade:** A transparência exige que os sistemas de IA sejam compreensíveis, inteligíveis e passíveis de auditoria (Brasil, Chile, Colômbia, Equador).
- **Transparência na Interação:** O usuário deve ser claramente informado de que está interagindo com um sistema de IA (identificação) e de como as decisões foram tomadas (Chile, Colômbia, Equador).
- **Obrigações de Alto Risco/Governança:** A transparência é intensificada para sistemas de maior risco, exigindo documentação detalhada, avaliações de impacto públicas e mecanismos de revisão humana (Brasil, Colômbia, Equador, Uruguai).

Figura 1 - Visão Macro sobre a Transparência da IA no Mercosul



MODELO PROPOSTO

Dada a relevância da transparência como um princípio central para os países do Mercosul, e considerando-se que o provimento de transparência da IA varia em função do caso de uso, da organização e do setor, existem estratégias que as empresas podem adotar ao conceber sistemas de IA (Jonker *et al.*, 2025). Tais estratégias incluem o estabelecimento de princípios claros de confiança e transparência, a sua implementação prática e a sua incorporação em todo o ciclo de vida da IA (Jonker *et al.*, 2025).

No caso de uso, o setor e o público-alvo do modelo deverão orientar as informações a serem divulgadas, sendo que sistemas de alto risco exigirão dados mais abrangentes do que sistemas de baixo risco (Jonker *et al.*, 2025). Cada etapa do ciclo de vida de desenvolvimento do sistema de IA pode contribuir com informações, distribuindo a responsabilidade por todo o ecossistema, em vez de atribuí-la a um único indivíduo. Plataformas de *software* e ferramentas estão disponíveis para auxiliar na automação da coleta de informações e em outras atividades de governança da IA (Jonker *et al.*, 2025). A seguir estão descritas as principais características para descrever sobre um modelo.

FORMATO E DOCUMENTAÇÃO

O formato da informação depende, adicionalmente, do público-alvo. Por exemplo, se a informação é destinada ao consumidor, ela deve ser facilmente compreensível; se for voltada para equipes técnicas (como cientistas de dados ou reguladores), exigirá um alto nível de detalhamento técnico. Os tipos de documentos a serem abrangidos podem incluir:

- Um documento vivo de conformidade.
- Páginas oficiais de políticas, detalhando como a organização implementa as iniciativas de transparência da IA.
- Recursos educacionais para auxiliar os usuários a compreenderem como a IA é empregada e como pode afetar a experiência do cliente.
- Atividades oficiais focadas na ética da IA na organização.
- Artigos de pesquisa e/ou outras comunicações para oferecer *insights* sobre o uso da IA na organização.

ÍTEM DE DIVULGAÇÃO DO MODELO

A divulgação do modelo pode abranger a totalidade ou parte das seguintes informações (Jonker *et al.*, 2025):

- Identificação: Nome do modelo, Propósito, Nível de risco, Política de modelo, Geração de modelos, Domínio pretendido, e Informações de contato.

- Desempenho e Qualidade: Dados de treinamento, Precisão de treinamento e teste, Viés, Métricas de robustez adversarial, Métricas de imparcialidade e Métricas de explicabilidade.

Esta lista é considerada abrangente e inclui itens de documentação frequentemente exigidos em regulamentações de Inteligência Artificial, sobretudo para sistemas de Alto Risco. Os itens serão detalhados a seguir, sendo mais explicitamente previstos na legislação, como o Projeto de Lei n.º 2.338/2023 (PL 2338/2023) do Brasil e em projetos similares alinhados ao modelo da União Europeia (como o do Equador), visto que estes exigem o estabelecimento de uma estrutura de governança e gestão de riscos baseada em documentação técnica e não técnica.

Por exemplo, a documentação de sistemas de inteligência artificial (Jonker *et al.*, 2025), especialmente daqueles classificados como de alto risco, exige a identificação clara e formal do **nome do modelo**, elemento essencial para fins de rastreabilidade, auditoria e registro oficial, conforme implícito no sistema de registro previsto no Projeto de Lei n.º 2338/2023. Associado a isso, o **propósito do modelo** deve ser explicitamente definido, abrangendo sua finalidade, o contexto de uso e as tarefas a que se destina, aspecto central para a determinação do nível de risco e para assegurar que o sistema seja empregado de acordo com o uso pretendido, nos termos do artigo 17 do referido projeto de lei (Brasil, 2023).

Outro elemento fundamental é a definição do **nível de risco** do sistema de IA (Jonker *et al.*, 2025), classificado em categorias como risco inaceitável, alto, limitado ou mínimo, com base no potencial impacto sobre direitos fundamentais e a segurança. Essa classificação é determinante para a aplicação das obrigações regulatórias correspondentes, conforme os artigos 17 e 19 do PL 2338/2023 (Brasil, 2023). Complementarmente, a **descrição da geração do modelo** (Jonker *et al.*, 2025) deve contemplar o processo de desenvolvimento, incluindo métodos de treinamento, arquitetura empregada e decisões tomadas ao longo do ciclo de vida do sistema, atendendo às exigências implícitas de documentação técnica e qualidade previstas para sistemas de alto risco.

A definição do **domínio pretendido** também é indispensável, especificando os setores, ambientes e contextos geográficos ou demográficos para os quais o modelo foi projetado e validado, de modo a prevenir usos indevidos, conforme a lógica de avaliação de risco baseada na finalidade e no contexto (Jonker *et al.*, 2025). No que se refere aos **dados de treinamento** (Jonker *et al.*, 2025), é necessária a apresentação de informações detalhadas sobre a origem, curadoria, representatividade, volume e metodologias empregadas no tratamento e mitigação de vieses, em consonância com o artigo 21, §1º, do PL 2338/2023 (Brasil, 2023).

A avaliação do desempenho do sistema deve incluir as **métricas de precisão de treinamento e teste** (Jonker *et al.*, 2025), como *accuracy*, *recall* e *F1-score*, a fim de verificar se o modelo atende aos padrões de desempenho estabelecidos, conforme as obrigações de gestão de risco e qualidade previstas nos artigos 19 e 21 (Brasil, 2023). De igual modo, é imprescindível a **análise de vieses** (Jonker *et al.*, 2025), com o registro dos potenciais vieses sistêmicos identificados nos dados ou no modelo, bem como das medidas adotadas para mitigá-los, em observância aos princípios de não discriminação, conforme o artigo 19, inciso III (Brasil, 2023).

No âmbito da segurança e confiabilidade, devem ser consideradas as **métricas de robustez adversarial** (Jonker *et al.*, 2025), que avaliam a resistência do sistema a ataques destinados a manipular suas saídas, em alinhamento com as exigências de segurança cibernética e gestão de riscos. Adicionalmente, as **métricas de imparcialidade** (Jonker *et al.*, 2025) permitem avaliar quantitativamente se o modelo produz resultados discriminatórios com base em atributos protegidos, complementando a análise de vieses e reforçando a proteção contra discriminação prevista no marco legal.

Por fim, a documentação deve contemplar **métricas de explicabilidade** (Jonker *et al.*, 2025), indicando o grau de transparência e interpretabilidade do modelo, inclusive por meio do uso de técnicas de inteligência artificial explicável (XAI), de modo a assegurar o cumprimento dos princípios de transparência e explicabilidade, bem como a realização

de avaliações de impacto algorítmico, conforme os artigos 2º, inciso VI, e 24 do PL 2338/2023 (Brasil, 2023). A inclusão de **informações de contato** do responsável pelo sistema de IA, como o controlador ou operador e o Encarregado de Dados, é igualmente necessária para garantir a comunicação com autoridades reguladoras e usuários afetados, em consonância com as exigências de documentação técnica e transparência ao público (Jonker *et al.*, 2025).

CONCLUSÃO

O presente estudo teve como objetivo central abordar a delegação de decisões a sistemas de Inteligência Artificial (IA) e a consequente discussão sobre a responsabilidade humana. Esta problemática é acentuada pela velocidade, pela complexidade da cadeia de desenvolvimento e uso (*multiple hands/things*), e, sobretudo, pela lacuna de conhecimento e explicabilidade inerente às operações da IA.

O desenvolvimento do trabalho incluiu, ademais, uma abordagem conceitual da transparência nas normativas dos países do Mercosul, identificando-a como um fator mitigador essencial para a redução de vieses. A análise da legislação sobre proteção de dados e IA na região culminou na proposição de uma visão macro dos requisitos de transparência regionais.

O resultado alcançado é um modelo de governança que delinea as principais tarefas requeridas para a IA Responsável no Mercosul. Este modelo considera os aspectos mais robustos das legislações nacionais e a sinergia regulatória entre os países do bloco. Em suma, o estudo oferece uma ferramenta valiosa para que desenvolvedores possam refletir, de forma contínua, sobre a importância da transparência na região e as melhores práticas para que modelos de IA criados em um país do Mercosul sejam empregados em outro, garantindo a mitigação de riscos éticos e a interoperabilidade regulatória.

REFERÊNCIAS

ARGENTINA. Poder Executivo Nacional. *Resolução n. 87/2023*. Projeto de Lei de Proteção de Dados Pessoais. [Mensagem]. Buenos Aires, [2023]. Disponível em:

https://www.argentina.gob.ar/sites/default/files/mensajeyproyecto_leydpd2023.pdf. Acesso em: 30 out. 2025.

BLANCHET, Atahualpa. *Inteligência artificial e direitos humanos*. [Aula aberta online]. Disponível em:

<https://www.instagram.com/p/DO9AbNmE67s/>. [S. l.]: [s. n.]. Acesso em: 29 set. 2025.

BRASIL. *Lei n. 13.709, de 14 de agosto de 2018*. Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília, DF: Senado Federal, [2024]. Disponível em:

https://www2.senado.leg.br/bdsf/bitstream/handle/id/658231/Lei_geral_protecao_dados_pessoais_led.pdf. Acesso em: 30 out. 2025.

BRASIL. Senado Federal. *Projeto de Lei n. 2.338, de 2023*: dispõe sobre o uso da inteligência artificial no Brasil. Brasília, DF: Senado Federal, [2025]. Disponível em:

<https://legis.senado.leg.br/sdleg-getter/documento?dm=9347622&ts=1742240889313&disposition=inline>. Acesso em: 30 out. 2025.

CHILE. Senado Federal. *Boletim n. 15.869-19 e 16.821-19*: regula os sistemas de inteligência artificial. [Projeto de Lei]. [S. l.], [2025]. Disponível em:

<https://www.camara.cl/verDoc.aspx?prmID=35352&prmTIPO=OFICIO>. Acesso em: 30 out. 2025.

COECKELBERGH, Mark. *Ética na inteligência artificial*. São Paulo: Ubu Editora, 2024.

COLÔMBIA. Governo Federal. *Lei Estatutária n. 1.581, de 17 de outubro de 2012*. Dispõe sobre a proteção de dados pessoais. Bogotá, [2023]. Disponível em:

<https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=49981>. Acesso em: 31 out. 2025.

COLÔMBIA. Ministério da Ciência, Tecnologia e Inovação. *Projeto de Lei*: por meio do qual se regula a inteligência artificial na Colômbia para garantir seu desenvolvimento ético e responsável e se ditam outras disposições. [2025]. Disponível em: https://minciencias.gov.co/sites/default/files/upload/noticias/pl_ia_finalizado.pdf. Acesso em: 31 out. 2025.

DA SILVA, Daniela A. *et al.* Principais características para o uso responsável da IA. In: CONFERÊNCIA LATINO-AMERICANA DE ÉTICA EM INTELIGÊNCIA ARTIFICIAL, 2024. São Paulo: SBC, 2024. p. 125–128.

DA SILVA, Daniela A.; MARQUES, J. Ethical considerations when using LLMs. In: AMERICAS CONFERENCE IN INFORMATION SYSTEMS, 2025.

DA SILVA, Daniela A.; MARQUES, Johnny C.; DA SILVA, Delmo M. Ética na inteligência artificial: Princípios para a tomada de decisão responsável. In: INTERNATIONAL WORKSHOP ETHICS AND AI — AIRES S2/E1 — RTAIM, 2025, Rio de Janeiro; Porto. Trabalho apresentado. Rio de Janeiro, Brasil: UFRJ; Porto, Portugal: UP, 2025.

DA SILVA, Daniela A.; MARQUES, Johnny C.; TASINAFFO, Paulo M. Mapping the asymmetries of graduate programs in brazil: modelling, visualization and reporting of estimates. *IEEE Revista Iberoamericana de Tecnologías del Aprendizaje*, v. 18, n. 3, 2023.

DA SILVA, Daniela A.; MARQUES, Johnny C.; TASINAFFO, Paulo M. Why machine learning bias is bad? In: WORKSHOP OF DATA SCIENCE— CEDS ITA, 2021. Trabalho apresentado. São José dos Campos, SP: ITA, 2021.

EQUADOR. Assembleia Nacional. Resolução n. [s.n.]. *Lei Orgânica de Regulação e Promoção da Inteligência Artificial no Equador*. [Projeto de Lei]. [2025]. Disponível em: <https://www.asambleanacional.gob.ec/es/multimedios-legislativos/97303-proyecto-de-ley-organica-de-regulacion> Acesso em: 31 out. 2025.

JONKER A. et al. *O que é a transparência de IA*. [S. l.]: IBM, 2025. Disponível em: <https://www.ibm.com/br-pt/think/topics/ai-transparency>. Acesso em: 29 out. 2025.

MERCOSUL. *Declaração de ministros e altas autoridades sobre direitos humanos dos Estados Partes do Mercosul sobre os princípios dos direitos humanos no âmbito da Inteligência Artificial*. [2023]. Disponível em: https://documentos.mercosur.int/simfiles/declaraciones/98118_AT_TXNW0C.docx. Acesso em: 27 out. 2025.

MÉXICO. Senado Federal. *Lei Federal de Proteção de Dados Pessoais em Posse de Particulares (LFPDPPP)*. Cidade do México, [2024a]. Disponível em: <https://www.gob.mx/indesol/documentos/ley-federal-de-proteccion-de-datos-personales-en-posesion-de-los-particulares>. Acesso em: 01 fev. 2026.

MÉXICO. Senado Federal. *Lei para a Regulação Ética da Inteligência Artificial para os Estados Unidos Mexicanos*. [Projeto de Lei]. Cidade do México, [2024b]. Disponível em: http://sil.gobernacion.gob.mx/Archivos/Documentos/2023/04/asun_4551867_20230420_1680209419.pdf. Acesso em: 31 out. 2025.

MILLER, Tim. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, v. 267, p. 1–38, 2019.

O'DONNELL, J. A IA deve nos lisonjear, nos corrigir ou apenas nos informar? *MIT Technology Review Brasil*, [2025]. Disponível em: <https://mittechreview.com.br/companion-reinforcing-behaviors-ia/>. Acesso em: 27 out. 2025.

PERU. Governo Federal. *Lei n. 31.814, de 24 de julho de 2023*. Lei que promove o uso da inteligência artificial em favor do desenvolvimento econômico e social do país. Lima, [2023]. Disponível em: <https://www.gob.pe/institucion/congreso-de-la-republica/normas-legales/4565760-31814> Acesso em: 31 out. 2025.

RICKMAN, S. Evaluating gender bias in large language models in long-term care. *BMC Medical Informatics and Decision Making*, v. 25, n. 1, p. 274, 2025.

SLATTERY, Philip *et al.* *The AI risk repository: A comprehensive meta-review, database, and taxonomy of risks from artificial intelligence*. [S. l.]: [s. n.], 2024. [Disponível em: [s. l.]. Acesso em: [s. d.].]

URUGUAI. Lei de Neurotecnologia e Proteção de Informação. *Lei n. 20.212*, art. 74. [Projeto de Lei sobre IA e Neurotecnologias]. [2025]. Disponível em: <https://www.impo.com.uy/bases/leyes/20212-2023>. Acesso em: 31 out. 2025.

YANG, Kai; QINAMI, Kenzo; FEI-FEI, Li; DENG, Jia; RUSSAKOVSKY, Olga. Towards fairer datasets: Filtering and balancing the distribution of the people subtree in the ImageNet hierarchy. In: CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY, 2020. p. 547–558.

FORMAÇÃO DOCENTE E IA: A EXPERIÊNCIA DA UNESP

FACULTY* DEVELOPMENT AND AI: THE UNESP EXPERIENCE

DOI: 10.29327/5798915.1-5

*José Alexandre Matelli*⁵
*Maurício César Delamaro*⁶

Resumo: Em meio aos incontáveis desafios que a massificação da IA, especialmente da IA generativa associada a grandes modelos de linguagem, impõem à educação superior, ter o corpo docente preparado para enfrentá-los se torna crucial. Tal preparação vai além da mera capacitação para operar diferentes modelos de IA: antes, envolve uma formação pedagógica robusta alicerçada nos fundamentos da educação e nas bases epistemológicas que sustentam as práticas de ensino superior, de tal modo que essa formação possibilite ao professor utilizar a IA com intencionalidade, com base em referenciais teóricos alinhados àquela intenção. Neste capítulo, pretendemos compartilhar a experiência da Unesp na formação profissional do seu corpo docente, atualmente amparada por uma política institucional que é um desdobramento dos esforços pioneiros do grupo da Prof^a Adriana Chaves, iniciados em 2006, culminan-

* O termo “*faculty*” se refere ao corpo docente e é mais comumente usado na América do Norte. Assim, “*faculty development*” seria o equivalente à nossa formação docente. Já no Reino Unido, Austrália e Europa, o termo “*academic development*” é mais utilizado, guardando o mesmo significado. Optamos aqui pelo termo “*faculty development*”, por parecer menos ambíguo do que “*academic development*”.

⁵ Universidade Estadual Paulista (UNESP), Coordenadoria de Desenvolvimento Profissional e Práticas Pedagógicas (CDeP3), jose.a.matelli@unesp.br.

⁶ Universidade Estadual Paulista (UNESP), Faculdade de Engenharia e Ciências, campus de Guaratinguetá, mauricio.delamaro@unesp.br.

do com a constituição da EFD – Escola de Formação dos Docentes da Unesp e no PROFOR – Programa de Formação Pedagógica e Desenvolvimento Profissional Docente. Recentemente, a EFD não ficou alheia ao advento da IA generativa e seus impactos na vida universitária, ofertando oficinas que tinham como tema o uso da IA na prática pedagógica, as implicações éticas e a política institucional de uso de IA na universidade. Por fim, apresentamos alguns dados para compreender como a IA e a formação docente em IA tem impactado a prática pedagógica dos docentes da Unesp.

Palavras-chave: Formação Docente; Prática Pedagógica; Política Institucional.

Abstract: Amid the numerous challenges imposed by AI massification in higher education, especially generative AI associated with large language models, having a well-prepared faculty is crucial. Such preparation should not be restricted to merely operate different AI models: rather, it should involve robust pedagogic training grounded in educational principles and the epistemological foundations that support the teaching practices in higher education. This way, the AI can be used with a particular intention, based on a theoretical framework aligned with that intention. In this chapter, we share the experience of São Paulo State University (Unesp) in faculty professional development, currently supported by an institutional policy originated from the pioneer efforts by the group led by Prof. Adriana Chaves in 2006. Those efforts culminated in the creation of the Faculty Development School (EFD) and in the Pedagogical Training and Professional Faculty Development Program (PROFOR). Recently, the EFD has addressed the AI challenge and its impact in the university, offering workshops to discuss the use of AI in pedagogical practice, ethical implications and the institutional policy regarding AI. We also present some data to better understand how generative AI and the faculty development in AI impact the pedagogical practices of Unesp faculty.

Keywords: Faculty Development; Pedagogical Practices; Institutional Policy.

INTRODUÇÃO

A discussão sobre a formação dos professores do ensino superior no Brasil é recente, pois só nos últimos anos do século XX essa formação passa a ser um tema mais amplamente pesquisado e chega ao século XXI ocupando considerável espaço nas investigações e publicações (Salvador; Sordi, 2024). Alguns artigos acadêmicos recentes abordam a formação docente para o ensino superior no Brasil, destacando ausência de exigência legal para formação didático-pedagógica específica e iniciativas institucionais pontuais como forma de contornar esse problema (Campos; Morais, 2024; Vieira; Vieira, 2023). Mais recentemente ainda, a massificação da IA generativa associada a grandes modelos de linguagem trouxe e segue trazendo significativa complexidade ao cenário da formação docente para o ensino superior. De fato, para que os professores tenham o domínio do resultado esperado para o processo de ensino-aprendizagem, torna-se fundamental que a formação docente contemple a análise crítica e a formação prática para o uso da IA na sala de aula (Durso, 2024).

Reconhecendo-se como parte desse contexto, existem duas iniciativas na Unesp, de naturezas e finalidades distintas, que envolvem a formação de docentes de ensino superior. A mais recente é o Grupo de Estudos e Pesquisas em Pedagogia Universitária (GEPPU), iniciativa de cunho acadêmico que, desde 2010, tem a pedagogia universitária como objeto e campo de produção de conhecimento. A outra compõe uma série de ações institucionais voltadas para a formação dos professores da universidade nos últimos vinte anos e que se consolidam atualmente na instituição da Coordenadoria de Desenvolvimento Profissional e Práticas Pedagógicas (CDeP3), responsável pelo planejamento e execução da política institucional de formação docente da universidade. É justamente no âmbito da CDeP3 que a Unesp vem promovendo ações institucionais de formação docente para uso de IA.

No presente trabalho, apresentamos um histórico da experiência da Unesp na formação do seu corpo docente a partir das iniciativas associadas principalmente à CDeP3, sem deixar de mencionar as atividades de pesquisa em pedagogia universitária do GEPPU. Apresentamos também a política institucional de formação dos docentes da Unesp e de que forma as questões relativas à IA foram incorporadas institucionalmente. Por fim, apresentamos alguns dados visando qualificar a compreensão de como a IA tem impactado a prática pedagógica dos professores da universidade.

A FORMAÇÃO DOCENTE NA UNESP

GRUPO DE ESTUDOS E PESQUISA EM PEDAGOGIA UNIVERSITÁRIA

O GEPPU é um grupo de pesquisa do Instituto de Biociências do campus de Rio Claro da Unesp, registrado no diretório de grupos de pesquisa⁷ do CNPq desde 2010 e certificado pela instituição. Segundo informações disponíveis no registro do grupo de pesquisa, as linhas de pesquisa do GEPPU estão organizadas em cinco grandes eixos, descritos a seguir, com informações obtidas da página⁸ do grupo.

Experiência, Diversidade e Linguagem no Contexto Universitário: Os estudos e pesquisas realizados neste eixo abordam a compreensão sobre os modos e os meios pelos quais se manifestam a Linguagem no âmbito das experiências universitárias. Entendendo as duas dimensões — Linguagem e Experiência — conjuntamente, recorre-se a elas para disparar e fomentar discussões maiores a respeito da formação humana, do pensamento crítico, das estruturas político-educacionais e/ou dos jogos de poder, no contexto universitário.

Organização do Trabalho Pedagógico Universitário: Os estudos e pesquisas realizados neste eixo abordam a identificação das diversas demandas que são exigidas dos docentes universitários, entre as quais se encontram a pesquisa, o ensino e a

⁷ Disponível em: <http://dgp.cnpq.br/dgp/espelhogrupo/1492729192922320>. Acesso em: 11 dez. 2025.

⁸ Disponível em <https://geppu.com.br/eixos>. Acesso em 11 dez. 2025.

extensão, buscando compreender como se dá o processo de construção da identidade docente e revelando as concepções que os professores do ensino superior trazem para o desempenho de sua profissão, a partir destes papéis acadêmicos.

Política, Gestão e Organização Institucional Universitária: Os estudos e pesquisas realizados neste eixo tratam da compreensão e análise dos impactos da intervenção política na gestão e organização institucional das instituições de ensino superior, discutindo os diferentes impactos na estrutura universitária, a partir de mudanças nos modelos de gestão. Somma-se a isso a intenção de compreender o processo de constituição identitária do gestor universitário.

Espaços e Processos de Formação na Universidade: Os estudos e pesquisas realizados neste eixo dizem respeito à Pedagogia Universitária dentro do contexto da formação pedagógica destinada ao docente universitário e à construção de espaços institucionais que se ocupam desta formação.

Fundamentos Epistemológicos da Pedagogia Universitária: Estudo dos fundamentos epistemológicos que orientam a constituição da Pedagogia Universitária considerando os contextos históricos, sociais, culturais, éticos e políticos nos quais este campo (conhecimento) foi e está sendo produzido nacional e internacionalmente, tomando a Educação Superior e suas contradições como realidade concreta de pesquisa. Os fundamentos epistemológicos são estudados em suas múltiplas articulações, buscando compreender contextos, sujeitos, referenciais teóricos e metodológicos acerca das investigações realizadas, a fim de desenvolver uma reflexão crítica e emancipadora sobre a origem e natureza da Pedagogia na materialidade da Educação Superior.

O GEPPU também organiza o CBPU – Congresso Brasileiro de Pedagogia Universitária. Com periodicidade bi-
enal, o objetivo do evento é fomentar as discussões e a produção de conhecimento acerca da complexidade epistemológica em torno da Pedagogia Universitária, ampliando a cientificidade deste campo de conhecimento.

COORDENADORIA DE DESENVOLVIMENTO PROFISSIONAL E PRÁTICAS PEDAGÓGICAS

A CDeP3 é a instância administrativa responsável por planejar e executar a política institucional de formação dos docentes da Unesp. A recente instituição da coordenadoria no final de 2022 (Unesp, 2022), porém, não reflete a já longa história da formação e aperfeiçoamento profissional dos professores da universidade, que teve origem em 2006 com o Programa de Oficinas de Estudos e Práticas Pedagógicas, vinculado à Pró-Reitoria de Graduação (Prograd). O Núcleo de Estudos e Práticas Pedagógicas (NEPP) foi criado em 2008 (Unesp, 2008). Também vinculado à Prograd, se institucionalizou formalmente a partir da consolidação do Programa de Oficinas de Estudos de Práticas Pedagógicas. O Centro de Estudos em Práticas Pedagógicas (CENEPP) foi criado em 2012 (Unesp, 2012). Não mais vinculado à Prograd e com mais autonomia, tinha o status administrativo de Unidade Complementar da Unesp, modalidade Instituto Especial, quando passou a se tornar um centro permanente, com sede administrativa no campus de Bauru, coordenação e comissões locais em todos os campi. O CENEPP era oficialmente denominado Centro de Estudos em Práticas Pedagógicas “Prof. Adriana Josefa Ferreira Chaves”, em homenagem a esta pioneira nas questões de formação docente na Unesp e militante da causa da educação. Em 2018, as tecnologias digitais convergiram para as ações do CENEPP pela incorporação da infraestrutura do Núcleo de Educação a Distância (NEaD), resultando na criação do Instituto de Estudos e Pesquisa em Práticas Pedagógicas “Prof. Adriana Chaves” (IEP3) (Unesp, 2018), com sede no bairro do Ipiranga, em São Paulo, no antigo prédio do Instituto de Artes da Unesp.

Instituída em 2022 a partir da extinção do IEP3 (Unesp, 2022), a Coordenadoria de Desenvolvimento Profissional e Práticas Pedagógicas — Professora Adriana Chaves (CDeP3) tem como objetivo central planejar, executar e avaliar políticas e programas voltados à formação pedagógica e ao desenvolvimento profissional dos docentes da Unesp. Vinculada à Vice-Reitoria, a CDeP3 articula ações integra-

das com as Pró-Reitorias para fortalecer a qualidade do ensino superior, promovendo o acolhimento de novos professores e incentivando práticas de equidade e educação inclusiva em toda a universidade.

As áreas de atuação da CDeP3 são divididas em quatro frentes principais. A Área I foca na formação pedagógica e no desenvolvimento profissional contínuo; a Área II concentra-se na Educação a Distância (EaD) e nas Tecnologias Digitais de Informação e Comunicação (TDIC), sendo responsável pelo suporte metodológico e pela gestão do Ambiente Virtual de Aprendizagem; a Área III dedica-se à pesquisa e à disseminação de conhecimento em pedagogia universitária; e a Área IV provê apoio e assessoria pedagógica, além da produção de materiais didáticos diversificados.

A estruturação da CDeP3 é descentralizada, materializando-se nas unidades universitárias por meio dos Núcleos Locais. Existem 34 núcleos instalados (um em cada unidade da Unesp), compostos por representantes docentes, técnico-administrativos e discentes. Esses grupos têm a missão de colaborar com o planejamento local, definir diretrizes de formação continuada aprovadas pelos colegiados e registrar as atividades pedagógicas realizadas em seus respectivos campi, garantindo que as políticas institucionais alcancem a realidade específica de cada unidade.

Para viabilizar a produção de conteúdo e o suporte às tecnologias educacionais, a CDeP3 conta com estúdios de gravação e infraestrutura tecnológica. Esses espaços padronizados são destinados para a criação de materiais audiovisuais, videoaulas e recursos didáticos interativos que apoiam tanto os cursos na modalidade a distância quanto as iniciativas de ensino híbrido e presencial. A gestão desses recursos permite que a coordenadoria ofereça suporte técnico e pedagógico especializado, consolidando a Unesp como referência na inovação das práticas de ensino e aprendizagem.

POLÍTICA INSTITUCIONAL DE FORMAÇÃO DOCENTE NA UNESP

A Resolução Unesp nº 23, de 17 de junho de 2025 (Unesp, 2025), estabelece as diretrizes fundamentais para a Po-

lítica Institucional de Formação Docente da Unesp, consolidando a atuação docente, em todas as suas dimensões, como uma atividade que demanda profissionalização e aperfeiçoamento contínuo. Para tanto, foi criada a Escola de Formação dos Docentes da Unesp (EFD), constituída como espaço de concepção, articulação e execução de ações de caráter institucional, tendo por objetivo geral propiciar desenvolvimento profissional e aprimoramento pedagógico continuado a docentes da Unesp, com eixos norteadores, linhas de ação, metas e indicadores de desempenho a serem definidos no Plano de Desenvolvimento Institucional⁹. A EFD é integrada administrativamente à CDeP3, especificamente à sua Área I de atuação.

No âmbito da EFD, institui-se o Programa de Formação Pedagógica e Desenvolvimento Profissional Docente (PROFOR). O PROFOR constitui o eixo central da política de formação, atuando como o instrumento executor das ações de capacitação e desenvolvimento profissional. A CDeP3 é a coordenadora do PROFOR no que diz respeito ao planejamento, à realização e ao registro das atividades. Do ponto de vista metodológico, o PROFOR se efetiva por meio de oficinas, cursos, seminários, discussões em comunidades de práticas e outras ações formativas presenciais e/ou a distância, fomentadas e organizadas pela EFD, abordando seis grandes dimensões:

I. Organização e gestão da estrutura acadêmico-administrativa da Instituição, incluindo atividades que buscam:

- a) promover, em conjunto com os departamentos de ensino, a ambientação do docente ingressante ao local e às condições de trabalho;
- b) promover a integração do docente ingressante com os demais servidores;
- c) propiciar ampla visão da organização e do funcionamento da Instituição;
- d) evidenciar a estrutura da instituição, seus objetivos, programas e planos;

⁹ Disponível em: <https://www2.unesp.br/portal#!pdi>. Acesso em: 18 dez. 2025.

II. Direitos humanos, equidade, diversidade e inclusão, com atividades formativas de caráter didático-pedagógico, contemplando os temas de:

- a) acessibilidade e inclusão educacional;
- b) direitos humanos;
- c) relações étnico-raciais e relações de gênero;
- d) liberdade e igualdade de oportunidades;
- e) reconhecimento comum e respeito;
- f) responsabilidade e justiça social;

III. Fundamentos educacionais e bases epistemológicas do ensino superior: atividades relacionadas aos fundamentos da educação e às bases epistemológicas que sustentam as práticas de ensino na Universidade;

IV. Pesquisa no ensino superior: atividades relacionadas à pesquisa, com o objetivo de promover compreensão aprofundada das diversas abordagens epistemológicas e metodológicas da ciência e possibilitar que os docentes integrem, em sua prática pedagógica, valorização crítica e contextualizada do conhecimento científico;

V. Atividades relacionadas à extensão no contexto universitário, orientadas para a promoção do impacto e da transformação sociais, que:

- a) realizem formação, objetivando direcionar as práticas docentes para ações de extensão e de pesquisa;
- b) promovam o impacto das ações na formação acadêmica;
- c) estabeleçam relação entre a Universidade e outros setores da sociedade;
- d) realizem a interação dialógica, a interdisciplinaridade, a interprofissionalidade e a indissociabilidade entre ensino, pesquisa e extensão;
- e) possibilitem a atuação do aluno como protagonista de sua formação técnica para obtenção de competências necessárias à atuação profissional e de sua formação cidadã;

f) discutam formas de fomento e de desdobramentos das ações extensionistas no âmbito de publicações e outros produtos acadêmicos;

VI. Metodologias, ferramentas e recursos educacionais do ensino superior: atividades que abordam elementos pedagógicos necessários ao exercício da docência no ensino superior, enfatizando o aprimoramento do processo ensino-aprendizagem nos cursos, programas e projetos oferecidos pela Unesp.

Entendendo a IA generativa tanto como uma ferramenta como um recurso educacional, é nessa última dimensão que ações voltadas a uma formação para uso da IA na prática pedagógica dos docentes foram propostas e executadas recentemente na Unesp.

IA E FORMAÇÃO DOCENTE NA UNESP

OFICINAS EM IA

No dia 30 de novembro de 2022, o ChatGPT foi disponibilizado ao público. Tomando este evento como o marco zero da massificação da IA generativa, é razoável admitir que a Unesp respondeu rapidamente a esse desafio, oferecendo sua primeira oficina em IA nos dias 3 e 4 de abril de 2024. O evento intitulado “1º Ciclo de Oficinas Pedagógicas em Tecnologias Digitais: o uso da Inteligência Artificial na prática docente” foi oferecido em parceria com a STHEM-Brasil (a parceria se estendeu para todos os ciclos realizados posteriormente). O objetivo principal foi capacitar o corpo docente para o uso crítico de algumas ferramentas de IA generativa, integrando-as aos processos de ensino e aprendizagem. Foram oferecidas no total 130 vagas a docentes da Unesp e a atividade foi realizada 100% no formato presencial, no auditório da CDeP3, bairro Ipiranga, na cidade de São Paulo. As atividades desenvolvidas na oficina são apresentadas em resumo no Quadro 1. O evento foi todo registrado em vídeo e está disponível¹⁰ na íntegra no canal da CDeP3.

¹⁰ Disponível em: https://www.youtube.com/playlist?list=PLJ3JXZ6O-seJC45h4prH4D_ksrFAMREa8C. Acesso em: 18 dez. 2025. Vídeos dos demais ciclos também estão disponíveis.

Quadro 1. Resumo das atividades desenvolvidas no “1º Ciclo de Oficinas Pedagógicas em Tecnologias Digitais: o uso da Inteligência Artificial na prática docente”

Tema	Motivação	Objetivo na educação
ChatGPT	Compreender o avanço da IA em linguagem natural	Integrar assistentes de IA para apoio educacional e desenvolvimento de habilidades de escrita
Edge AI	Entender a IA aplicada em dispositivos móveis	Utilizar tecnologias de ponta para personalizar e aprimorar a aprendizagem
Magic school	Explorar ambientes virtuais de aprendizagem	Criar experiências imersivas para melhor engajamento e compreensão dos alunos
Gemini	Conhecer sistemas de IA para conteúdo educacional	Implementar ferramentas de IA para criar conteúdos didáticos interativos
DALL-E	Aprender sobre IA em arte e design gráfico	Desenvolver materiais visuais atrativos e educacionais com IA
Revisão e discussão	Refletir sobre o aprendizado e possíveis aplicações	Planejar a integração das tecnologias de IA na prática pedagógica

No mesmo ano de 2024, nos dias 3 e 4 de setembro, foi oferecido o “2º Ciclo de Oficinas Pedagógicas em Tecnologias Digitais: o uso da Inteligência Artificial na prática docente”, com a mesma programação do evento anterior. Novamente, foram ofertadas 130 vagas a docentes da Unesp, com atividades 100% no formato presencial realizadas no auditório da CDeP3, no Ipiranga.

Durante a realização das duas oficinas de 2024, percebeu-se que os docentes traziam, com frequência, questões relacionadas aos dilemas éticos implicados na utilização da IA, tais como autoria, plágio e até mesmo a obsolescência da figura do professor. Essas questões eram abordadas concomitantemente pelo Comitê Superior de Tecnologia da Informação (CSTI), vinculado à Coordenadoria de Tecnologia da Informação da Unesp (CTInf). O CSTI elaborou uma proposta de política institucional sobre uso de IA na Unesp, que culminou com a aprovação, pelo Conselho de Ensino, Pesquisa e Extensão, da Resolução Unesp nº 13, de 22 de abril de 2025 (Unesp, 2025a), instrumento que oficializa essa po-

lítica em toda a universidade, com destaque para os princípios de uso (Art. 3º), as diretrizes gerais de uso (Art. 4º) e o uso nas atividades acadêmicas e administrativas (Art. 6º).

Acrescentando o surgimento de novas ferramentas a esse contexto, os ciclos realizados em 2025 tiveram suas programações atualizadas. No Quadro 2, a programação do “3º Ciclo de Oficinas Pedagógicas em Tecnologias Digitais: o uso da Inteligência Artificial na prática docente” incluiu um período para destacar a política de uso de IA na Unesp, seguido de uma mesa redonda sobre as questões éticas envolvidas no uso da IA na universidade. Para o quarto ciclo, a programação incluiu um período para atividades práticas, atendendo uma demanda dos docentes. Foi também dedicado um período para apresentar as IAs institucionais desenvolvidas pela CTInf, no âmbito do Laboratório do Futuro. Essas IAs institucionais são:

legIA: chatbot para tirar dúvidas sobre leis e normas da Unesp.

sapientIA: assistente para responder a perguntas sobre trabalhos científicos do Repositório Institucional Unesp.

pitIA: acesso a diferentes motores de IA generativa sem que os dados informados sejam usados para treinamento.

Quadro 2. Resumo das atividades desenvolvidas no “3º Ciclo de Oficinas Pedagógicas em Tecnologias Digitais: o uso da Inteligência Artificial na prática docente”

Tema	Motivação	Objetivo na educação
ChatGPT	Compreender o avanço da IA em linguagem natural	Integrar assistentes de IA para apoio educacional e desenvolvimento de habilidades de escrita
DeepSeek Qwen Chat	Ferramentas específicas para insights educacionais e suporte personalizado em tempo real	Utilizar tecnologias de ponta para personalizar e aprimorar a aprendizagem
Poe Gemini Perplexity	Explorar IAs generativas para fins educacionais	IAs para facilitar o aprendizado, oferecer suporte personalizado e fornecer respostas rápidas e precisas para dúvidas, promovendo uma experiência educacional mais eficiente e acessível
Políticas de uso de IA na Unesp	Apresentação da Resolução que trata sobre a utilização da Inteligência Artificial na Unesp.	
Mesa redonda	Questões éticas no uso de IA no ambiente acadêmico	

Quadro 3. Resumo das atividades desenvolvidas no “4º Ciclo de Oficinas Pedagógicas em Tecnologias Digitais: o uso da Inteligência Artificial na prática docente”

Tema	Motivação	Objetivo na educação
ChatGPT Gemini Copilot	Compreender o avanço da IA em linguagem natural	Integrar assistentes de IA para apoio educacional e desenvolvimento de habilidades de escrita
DeepSeek Napkin Perplexity	Explorar IAs generativas para fins educacionais	IAs para facilitar o aprendizado, oferecer suporte personalizado e fornecer respostas rápidas e precisas para dúvidas, promovendo uma experiência educacional mais eficiente e acessível
legIA sapienIA pitIA	Apresentação das IAs institucionais da Unesp	
Políticas de uso de IA na Unesp	Apresentação da Resolução que trata sobre a utilização da Inteligência Artificial na Unesp.	
Mesa redonda	Questões éticas no uso de IA no ambiente acadêmico	

IMPACTOS DOS CICLOS DE OFICINAS PEDAGÓGICAS NA ADOÇÃO DA IA

O número de professores inscritos e que participaram efetivamente dos ciclos é apresentado na Tabela 1. O total de participantes foi de 446 docentes, o que corresponde a aproximadamente 15% do quadro docente da Unesp. Para avaliar o impacto das OEPs na adoção da IA pelos professores participantes, foi elaborado um questionário¹¹ para levantar alguns dados sobre a adoção da IA na prática pedagógica, com especial atenção na dimensão do ensino de graduação e pós-graduação.

Até o momento da publicação deste capítulo, foram obtidas 42 respostas. Embora o tamanho dessa amostra não permita fazer inferências estatisticamente válidas dentro dos parâmetros padronizados de confiança e margem de erro, é possível extrair algumas informações de caráter qualitativo e orientativo.

¹¹ Por se enquadrar no disposto no Art. 1º, Parágrafo único, Inciso VII da Resolução CNS nº 510, de 7 de abril de 2016, a pesquisa não foi registrada no sistema CEP/CONEP.

Tabela 1. Participantes dos Ciclos de Oficinas Pedagógicas em Tecnologias Digitais: o uso da Inteligência Artificial na prática docente

Ciclo	Ano/Semestre	Inscritos	Participantes
1	2024/1	204	130
2	2024/2	126	120
3	2025/1	127	122
4	2025/2	80	74
	Total	537	446

Das 34 unidades universitárias da Unesp, responderam ao questionário docentes de 25 delas. A maior parte dos participantes é de docentes recém-contratados (Figura 1), com menos de cinco anos de casa, o que é esperado, dado que as gerações mais velhas tendem a ser mais restritivas em disponibilizar dados pessoais e ter menor interesse em IA (Kubovics, 2025), enquanto indivíduos mais jovens mostram mais opiniões positivas a respeito de robôs e IA (Méndez-Suárez; Monfort; Hervas-Oliver, 2023). Ainda que a proporção de novos docentes seja maior, é também significativa a presença de docentes mais experientes, com distribuição razoavelmente equilibrada a partir da faixa de 11 anos de exercício profissional. Nesse sentido, é particularmente interessante notar que o segundo grupo mais numeroso é o de docentes com mais de 26 anos de carreira.

Com relação à grande área de atuação (Figura 2), era mesmo esperada uma grande participação (38,1%) de docentes das exatas (Engenharias, Ciências Agrárias e Ciências Exatas e da Terra), já que a IA é um produto gerado com conhecimento dessas áreas. O contingente de docentes das grandes áreas que envolvem a biologia (Ciências da Saúde, Ciências Biológicas e Ciências Agrárias) é de 33,1%, tendo em mente essas grandes áreas concentram o maior número de docentes da Unesp. Embora em proporção menor (28,6%), o grupo das humanidades (Ciências Sociais Aplicadas; Linguística, Letras e Artes; Ciências Humanas) teve uma participação significativa, ainda mais sabendo-se que essas áreas concentram o menor número de docentes da universidade.

Figura 1. Tempo de casa dos respondentes

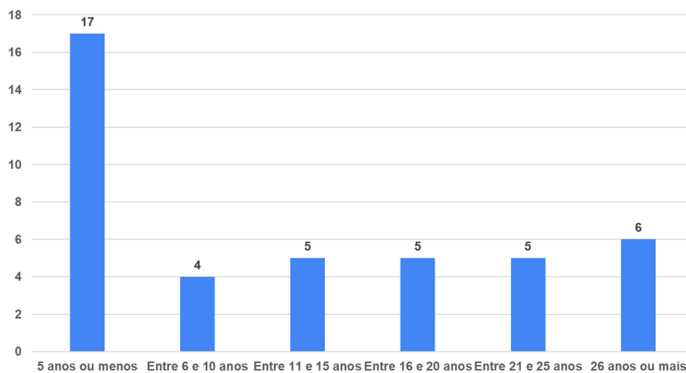
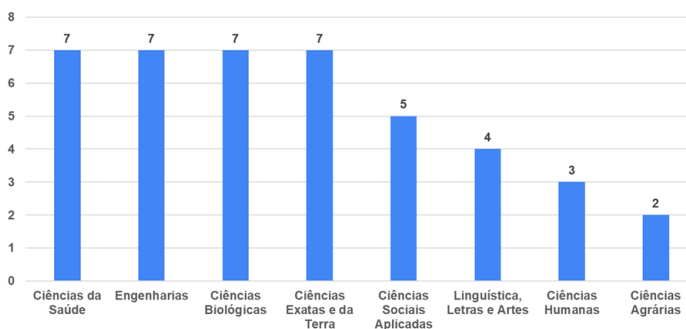


Figura 2. Grande área do respondente



Visando compreender o impacto das oficinas de IA na prática docente e obter subsídios para decisões estratégicas a respeito do futuro das oficinas, o questionário apresentou um conjunto de questionamentos agrupados nos seguintes temas:

- Efeitos da participação nos Ciclos de Oficinas Pedagógicas na prática pedagógica;
- Possíveis receios sobre o uso de IA na prática docente;
- Ações e políticas institucionais desejáveis relacionadas ao uso de IA na Unesp;
- Experiência com o uso de IA na prática pedagógica após a participação nos ciclos;
- Uso de IA nas diferentes dimensões da atuação universitária.

Para mensurar as atitudes, percepções e comportamentos dos respondentes em relação aos temas anteriores, foi adotada a Escala Likert, uma das ferramentas mais utilizadas em pesquisas de opinião, marketing e ciências sociais. O método permite que o respondente expresse o seu nível de concordância ou discordância em relação a uma afirmação específica. O nível de concordância e discordância é pontuado de acordo com a escala apresentada na Tabela 2.

O score de uma pergunta i qualquer é expresso pelo respectivo número de respondentes que respondeu “concordo plenamente” (n_{cpl}) ou “concordo parcialmente” (n_{cpr}), computado de acordo com a Equação 1. O score máximo da pergunta i é expresso como se todos os respondentes do questionário (n_r) tivessem respondido “concordo plenamente” com a questão i , isto é, $n_{cpl,i} = n_r$ (Equação 2). A razão de concordância r_i (Equação 3) é uma medida normalizada do nível de concordância: quanto mais próxima de 1 (ou de 100%), maior o nível de concordância com a questão i .

$$S_i = 4n_{cpl,i} + 3n_{cpr,i} \quad (1)$$

$$S_{m\acute{a}x,i} = 4n_r \quad (2)$$

$$r_i = S_i / S_{m\acute{a}x,i} \quad (3)$$

Tabela 2. Pontuação da Escala Likert para as respostas.

Resposta	Pontuação
Concordo plenamente	4
Concordo parcialmente	3
Nem concordo, nem discordo	2
Discordo parcialmente	1
Discordo totalmente	0

Para avaliar os efeitos da participação nos Ciclos de Oficinas Pedagógicas na prática pedagógica, foram apresentadas nove afirmações, conforme Quadro 4.

Quadro 4. Afirmações do bloco “Efeitos da participação nos Ciclos de Oficinas Pedagógicas na prática pedagógica”

Nº	Afirmiação
1	As oficinas ampliaram minha compreensão sobre o funcionamento e as limitações da IA
2	A experiência estimulou reflexões éticas e pedagógicas sobre o uso da IA
3	A participação demonstrou possibilidades concretas de uso da IA no processo ensino-aprendizagem
4	As oficinas contribuíram para quebrar medos e preconceitos em relação à IA
5	Passsei a perceber potencial da IA para reduzir tarefas repetitivas e automatizáveis
6	As oficinas fortaleceram meu domínio técnico sobre o uso de ferramentas de IA
7	As oficinas aumentaram minha confiança para aplicar IA na prática docente
8	Notei mudanças efetivas na minha prática pedagógica após as oficinas
9	Sinto-me mais amparado(a) por políticas ou ações institucionais sobre o tema

Figura 3. Avaliação sobre os efeitos da participação nos ciclos

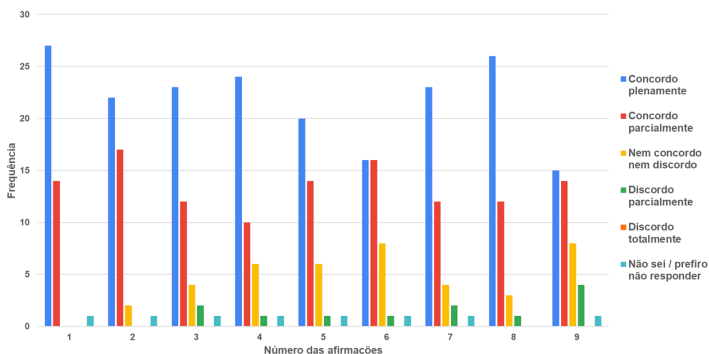
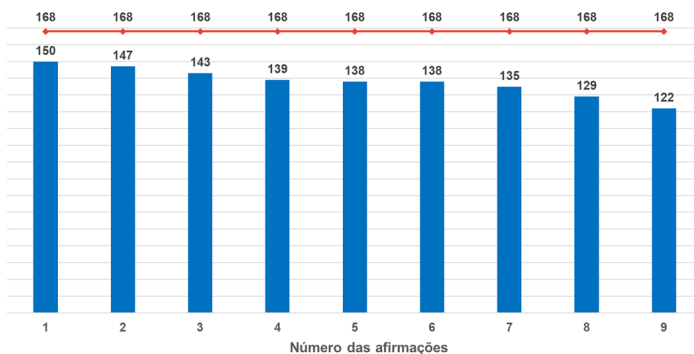


Figura 4. Efeitos da participação: scores de cada afirmação, em ordem decrescente.



Os níveis de concordância dos respondentes aparecem na Figura 3 e os *scores* calculados pela escala Likert, na Figura 4, em ordem decrescente. Nessa figura, a linha vermelha representa os *scores* máximos que seriam obtidos caso todos os respondentes tivessem expressado concordância plena.

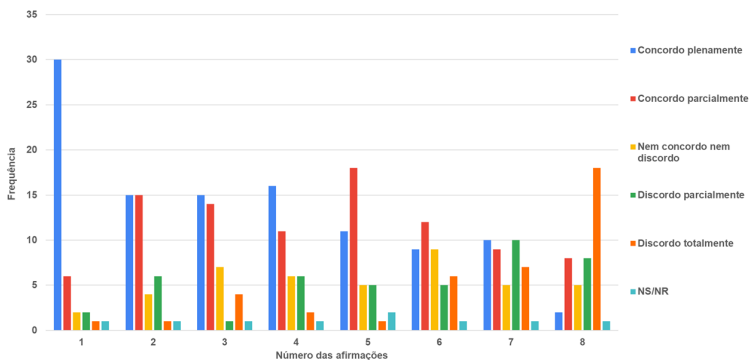
Quanto à eficácia da participação, para os fins do presente trabalho, admite-se que os Ciclos de Oficinas Pedagógicas em IA foram bem sucedidos no aspecto refletido pela pergunta i se $r_i \geq 80\%$. A partir da Figura 4, pode-se verificar que as oficinas foram efetivas, com 7 das 9 questões apresentando $r_i \geq 80\%$. Para os ciclos futuros, deveria ser dada mais atenção à identificação das barreiras que impedem a percepção de mudanças efetivas na prática pedagógica dos participantes, bem como estratégias para superá-las. Questões referentes a políticas e ações institucionais sobre uso de IA também devem ser mais bem trabalhadas, não só nas oficinas futuras, mas como um esforço institucional de divulgação e formação das políticas e ações, dado que elas existem.

Já os possíveis receios sobre o uso de IA na prática docente foram avaliados a partir de oito afirmações (Quadro 5), cujas concordâncias estão ilustradas na Figura 5.

Quadro 5. Afirmações do bloco “*Possíveis receios sobre o uso de IA na prática docente*”

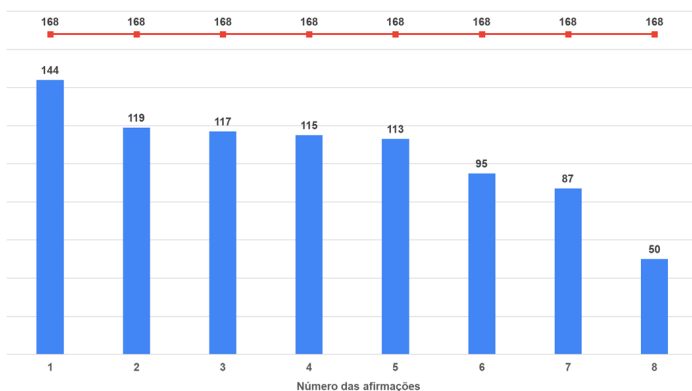
Nº	Afirmação
1	Preocupo-me com a possibilidade de plágio ou uso indevido de IA pelos alunos
2	Tenho dúvidas sobre a confiabilidade das respostas produzidas por IA
3	Sinto preocupação com a falta de critérios éticos e institucionais para o uso da IA
4	Temo que os estudantes passem a depender excessivamente das ferramentas de IA
5	Receio que o uso da IA dificulte a avaliação autêntica da aprendizagem
6	Tenho receio de não dominar suficientemente as ferramentas de IA para aplicá-las de modo adequado
7	Penso que o uso crescente de IA pode desumanizar as relações de ensino e aprendizagem
8	Tenho receio de que a IA reduza o papel do professor no processo de ensino

Figura 5. Respostas sobre os possíveis receios do uso da IA



Na Figura 6, os *scores* de cada uma das perguntas estão dispostos em ordem decrescente, juntamente da linha vermelha que representa os scores máximos, obtidos se todos os respondentes tivessem concordância plena com as afirmações.

Figura 6. Possíveis receios: scores de cada afirmação, em ordem crescente.



Dada a natureza das questões no caso dos receios, o critério da razão de concordância aqui se inverte: quanto menor o r_i , menor o receio e a oficina foi efetiva se $r_i \leq 20\%$. Assim, a ordem apresentada na Figura 7 é crescente. Sem muita surpresa, os receios são dominantes entre os participantes das oficinas, refletido pelo fato de todas as questões apresentarem $r_i \gg 20\%$. Em outras palavras, as

oficinas não foram efetivas na superação dos receios dos participantes. De longe, o maior receio é o plágio, o que de alguma forma se reflete também na preocupação com a falta de critérios éticos e institucionais para o uso da IA. Destacam-se também as dúvidas sobre a confiabilidade das respostas produzidas por IA e a possível dependência excessiva dos alunos das ferramentas de IA. Surpreende também que o menor receio, por longa margem, seja o da IA reduzir o papel do professor no processo de ensino. Diversas interpretações para essa percepção poderiam ser apresentadas, mas isso estaria fora do escopo deste trabalho e poderia ser mais bem explorado em um trabalho dedicado a essa investigação. Possíveis ações futuras para mitigar esses receios devem envolver um aprofundamento do caráter contínuo da formação docente para IA, bem como o estabelecimento de uma cultura organizacional que estimule o uso racional e ético da IA na prática pedagógica, inclusive com investimento em ferramentas e infraestrutura.

As percepções dos participantes das oficinas sobre as ações e políticas institucionais desejáveis relacionadas ao uso de IA na Unesp foram avaliadas através de oito questões (Figura 7), cujos *scores* estão ordenados conforme apresentado na Figura 8.

Figura 7. Respostas sobre ações e políticas institucionais para uso da IA

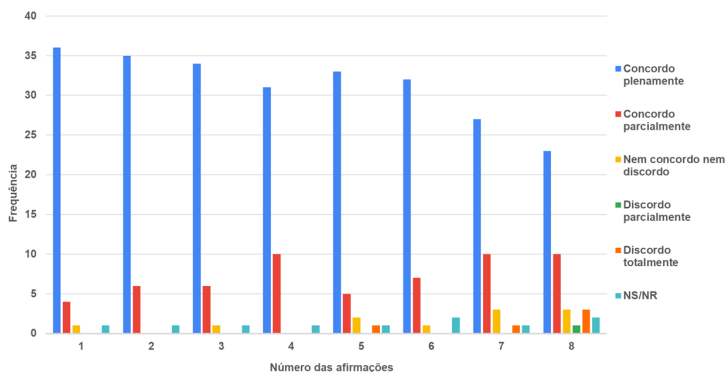
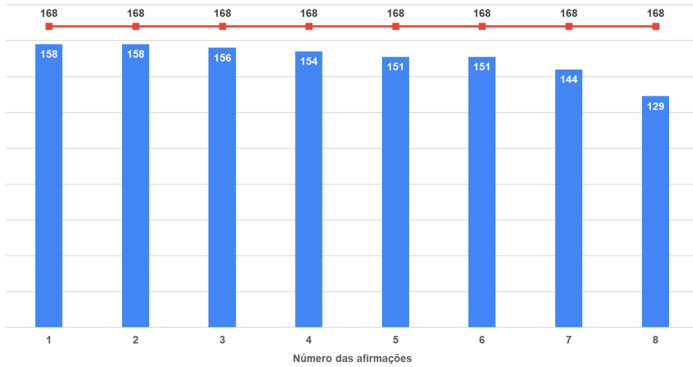


Figura 8. Ações e políticas institucionais: *scores* de cada pergunta, em ordem decrescente.



No que diz respeito ao tema de políticas institucionais, a concordância dos participantes com as questões apresentadas é muito expressiva, com 6 questões com $r_i \geq 90\%$ e apenas uma com $r_i < 80\%$. Algumas dessas percepções dialogam diretamente com alguns aspectos anteriormente discutidos, conforme sumarizado no Quadro 6 (outras relações poderiam ser exploradas, mas o texto ficaria demasiado extenso). A relação mostrada no Quadro 6 deixa evidente que é necessário mais esforço institucional para não só divulgar as políticas institucionais existentes, mas, mais importante, aprofundar o entendimento dessas políticas junto à comunidade universitária.

Quadro 6. Relação entre algumas percepções em diferentes temas

Tema	Percepção	Relação
Efeitos da participação nas oficinas	Sinto-me mais amparado(a) por políticas ou ações institucionais sobre o tema.	A falta de critérios éticos e institucionais é um receio fortemente percebido entre os participantes, que entendem que orientações mais claras nesse sentido deveriam estar disponíveis. Por outro lado, as oficinas não foram eficazes em mostrar que existem políticas institucionais, já que percepção do amparo institucionais não é tão forte.
Possíveis receios	Sinto preocupação com a falta de critérios éticos e institucionais para o uso da IA.	
Ações e políticas institucionais	Deveria haver orientações institucionais claras sobre o uso ético da IA em sala de aula.	

Para avaliar a experiência com o uso de IA na prática pedagógica após a participação nos ciclos, foram apresentadas oito perguntas (Quadro 7), respondidas conforme mostrado na Figura 9, sendo que os *scores* (Tabela 3) de cada uma das perguntas são apresentados em ordem decrescente na Figura 10. A linha vermelha nessa figura representa os scores máximos.

Quadro 7. Afirmações do bloco “Experiência com o uso de IA na prática pedagógica após a participação nos ciclos”

Nº	Afirmação
1	A IA tem contribuído para o planejamento das aulas
2	A IA tem estimulado novas práticas pedagógicas
3	A IA tem ajudado a otimizar o tempo de preparação de materiais
4	A IA tem ampliado minha reflexão sobre ética e autoria no ensino
5	Eu me sinto confortável e confiante para usar IA em sala de aula
6	O uso da IA tem aumentado o engajamento dos estudantes
7	O uso da IA tem melhorado a qualidade das avaliações
8	A IA tem favorecido a personalização da aprendizagem

Figura 9. Respostas sobre a experiência prática após participação nos ciclos

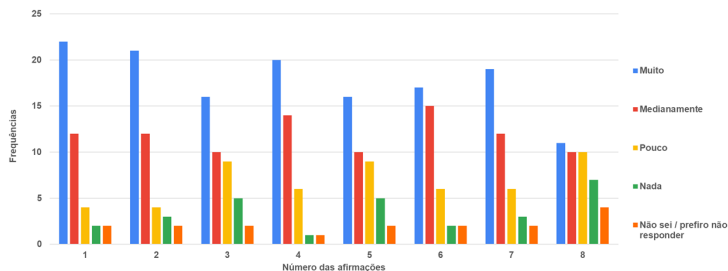
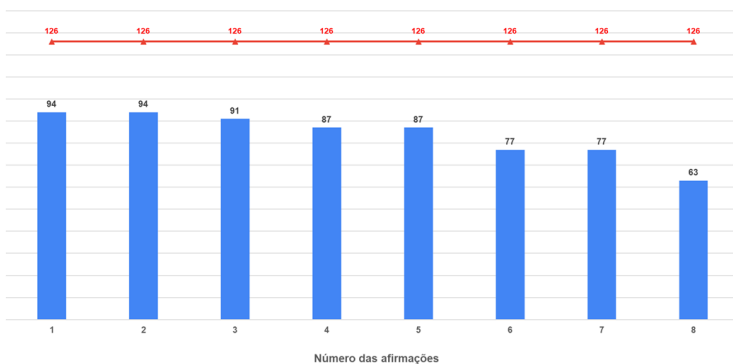


Tabela 3. Pontuação da Escala Likert para as respostas referentes à experiência após participação nas oficinas

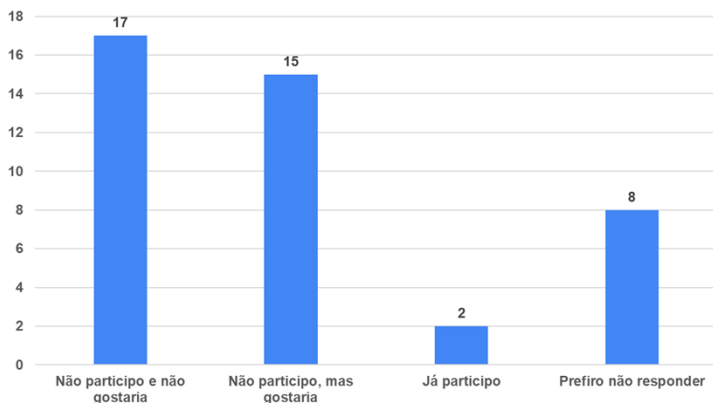
Resposta	Pontuação
Muito	3
Medianamente	2
Pouco	1
Nada	0

Figura 10. Experiência prática após participação: *scores* de cada pergunta, em ordem decrescente.



Todos os aspectos abordados na Figura 10 apresentaram $r_i < 80\%$, o que indica que, apesar da participação nas oficinas, a adoção da IA na prática docente ainda é um tanto tímida. Isso é um forte indicador de que a formação docente em geral, e em IA em particular, deve ter caráter continuado, com estímulo à troca de experiências. Isto é corroborado pelas perguntas “A Unesp deveria oferecer formação continuada sobre IA aplicada à prática pedagógica do(a) docente?”, que teve uma razão de concordância $r_i = 92,9\%$, e “É importante criar espaços de troca e colaboração entre docentes sobre práticas com IA?”, com $r_i = 91,7\%$ (Figura 9). Importante também que as atividades de formação sejam feitas nas Unidades Universitárias por meio dos Núcleos Locais, por estarem mais próximos do cotidiano, das especificidades e da cultura organizacional daquela comunidade, sem que seja ignorada a natureza das grandes áreas trabalhadas na Unidade. Nesse sentido, é fundamental que os participantes das oficinas possam participar como disseminadores em seus respectivos Núcleos Locais. No entanto, quando perguntados se gostariam de participar como disseminadores nos Núcleos Locais, a maior parte dos respondentes disse que não participa, nem gostaria de participar (Figura 11), o que faz com que ações de engajamento sejam consideradas ou que outras estratégias de formação continuada em IA nas Unidades precisem ser pensadas.

Figura 11. Disposição em atuar como disseminador no Núcleo Local.



Pode contribuir para a pouca adoção da IA a indisponibilidade das principais ferramentas. De fato, na Figura 9, a pergunta “A Unesp deveria disponibilizar ferramentas de IA seguras e acessíveis para uso docente?” teve uma razão de concordância $r_i = 89,8\%$. A Figura 12 mostra as ferramentas mais utilizadas pelos respondentes (poderiam indicar até três ferramentas), enquanto a Figura 13 mostra que quase um quarto dos respondentes pagam alguma ferramenta por conta própria.

Figura 12. Ferramentas mais usadas

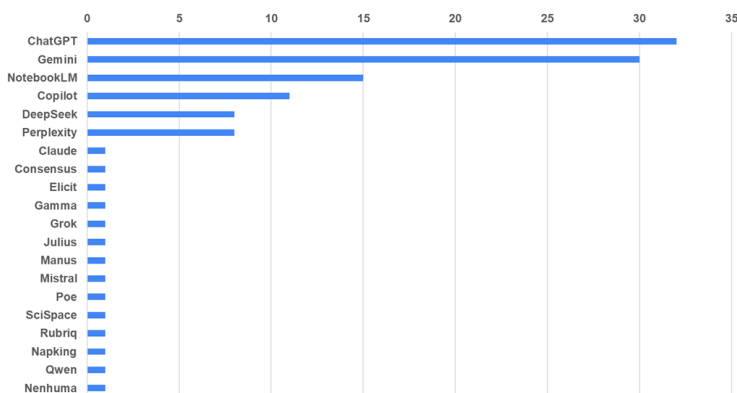
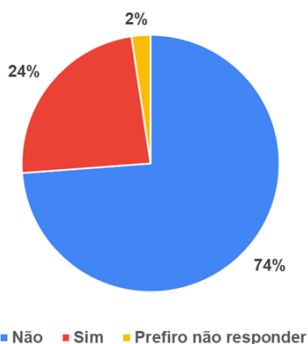
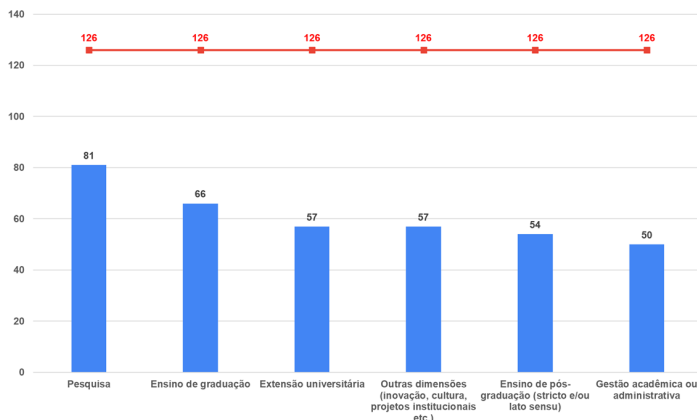


Figura 13. Docentes que pagam alguma ferramenta



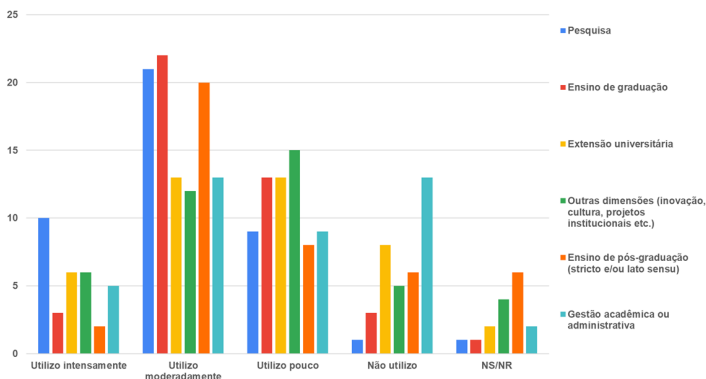
Se há certa timidez no uso da IA na prática pedagógica (isto é, na dimensão de ensino de graduação e de pós-graduação), a dimensão da pesquisa é a que apresenta maior score de utilização de IA¹², como observado na Figura 14, obtida a partir das respostas mostradas na Figura 15. Ainda assim, a percepção dessa utilização não é tão expressiva, apresentando $r_i = 64,3\%$. Os resultados da Figura 14 são indícios fortes de que a formação docente deve contemplar também as demais dimensões da atuação docente (extensão, gestão, inovação etc.).

Figura 14. Scores de uso de IA nas dimensões de atuação docente



¹² Nesse caso, a pontuação da Escala Likert foi: utilizo intensamente (3); utilizo moderadamente (2); utilizo pouco (1); não utilizo (0); não sabe/prefiro não responder não conta.

Figura 15. Resposta sobre uso de IA nas dimensões de atuação



TESTES ESTATÍSTICOS DE ALGUMAS HIPÓTESES

Nesta seção, são propostos dois testes estatísticos para avaliar duas hipóteses referentes aos receios e à utilização da IA pelas diferentes áreas de conhecimento. Os testes são o exato de Fisher (Freeman–Halton, 3×2) e o de independência pelo χ^2 .

RECEIOS E PREOCUPAÇÕES SOBRE USO DE IA POR ÁREA DE CONHECIMENTO

A aplicação dos testes de Fisher e de independência ($\alpha = 0,05$; tabelas 3×2) para avaliar os receios e preocupações sobre uso da IA nas diferentes áreas de conhecimento requereu o agrupamento de áreas. Ciências Agrárias; Ciências Exatas e da Terra e Engenharias foram agrupadas em Exatas. Linguística, Letras e Artes; Ciências Sociais Aplicadas e Ciências Humanas foram agrupadas em Humanas. Ciências da Saúde e Ciências Biológicas foram agrupadas em Biológicas. Devido ao tamanho da amostra, as respostas também precisaram ser agrupadas. As opções “concorda plenamente” e “concorda parcialmente” foram agrupadas em “concordância”. As demais opções foram agrupadas em “não concordância”.

Para ambos os testes, o que se busca é identificar se a concordância ou não com as afirmações depende ou está associada às áreas de conhecimento. Em outras palavras, a hipótese nula, H_0 , sustenta que a concordância com cada uma das afirmações não depende da área do conhecimento

do docente. Utilizou-se 5% como nível de significância do teste, sendo que as tabelas cruzadas tiveram sempre a estrutura 3 linhas (áreas) e 2 colunas (concordância ou não).

O p-valor é uma métrica de aceitação de H_0 , nos testes. Se p-valor $> 0,05$, H_0 é aceita; do contrário, é rejeitada. Quando H_0 é rejeitada, isso indica que a probabilidade de concordar com a afirmação varia conforme a grande área do conhecimento. Ou seja, não é apenas variação aleatória: a área em que o docente atua ajuda a explicar sua posição sobre aquela afirmação específica. Quando H_0 é aceita, as diferenças observadas entre áreas são compatíveis com flutuações amostrais e a concordância com a afirmação pode ser interpretada como amplamente compartilhado entre áreas. Os resultados são apresentados na Tabela 4.

Tabela 4. Medos e receios: resultados dos testes

Afirmação	Teste de Fischer		Teste do χ^2	
	p-valor	H_0 é aceita?	p-valor	H_0 é aceita?
1. A IA pode reduzir o papel do professor no processo de ensino	0,13	Sim	0,12	Sim
2. A IA pode gerar dependência excessiva dos estudantes	0,094	Sim	0,089	Sim
3. Falta de critérios éticos e institucionais claros para uso da IA	0,011	Não	0,016	Não
4. As respostas da IA não são suficientemente confiáveis	0,019	Não	0,031	Não
5. IA facilita plágio ou uso indevido pelos alunos	0,28	Sim	0,24	Sim
6. A IA dificulta a avaliação autêntica da aprendizagem	0,38	Sim	0,35	Sim
7. Receio de não dominar suficientemente as ferramentas de IA	0,052	Não (limitrofe)	0,039	Não
8. A IA pode desumanizar as relações de ensino e aprendizagem	0,15	Sim	0,15	Sim

Os resultados do teste de independência pelo χ^2 e do teste exato de Fisher mostraram-se consistentes. As inferências devem ser interpretadas considerando o tamanho amostral reduzido e o uso de reagrupamentos conceitualmente orientados, adotados para garantir a validade estatística das análises.

Tendo esses cuidados em mente, a interpretação dos testes permite as seguintes inferências:

- a. Em 3 das 8 questões analisadas observou-se evidência de associação entre a concordância dos docentes e a

grande área do conhecimento. Em duas delas, a associação foi confirmada de forma consistente por ambos os testes; em uma terceira, o resultado foi limítrofe, sendo indicado pelo teste do χ^2 , ainda que não confirmado pelo teste exato de Fisher, o que justifica sua consideração como indício relevante, embora com cautela interpretativa;

- b. As diferenças entre áreas existem, mas não são generalizadas, estando restritas a um subconjunto específico de receios. Isso sugere que a percepção dos riscos associados ao uso da IA não é amplamente determinada pela área de atuação, mas se manifesta de forma diferenciada apenas em determinadas dimensões;
- c. Os docentes das Ciências Humanas tendem a expressar maior sensibilidade a questões éticas e institucionais relacionadas ao uso da IA, bem como maior preocupação com a confiabilidade das respostas produzidas por essas ferramentas, dimensões nas quais a associação com a grande área foi estatisticamente confirmada;
- d. No que se refere ao domínio técnico das ferramentas de IA, observa-se um resultado limítrofe, que sugere menor receio entre docentes das Ciências Exatas, possivelmente associado a maior familiaridade instrumental. No entanto, esse achado deve ser interpretado com prudência, dado que a evidência estatística não é plenamente convergente entre os testes;
- e. Alguns receios mostram-se transversais às diferentes áreas, sem evidência estatística de associação, como preocupações com plágio e uso indevido da IA, dificuldades na avaliação autêntica da aprendizagem e desumanização do processo de ensino-aprendizagem, indicando que tais inquietações são amplamente compartilhadas entre os docentes, independentemente da área;
- f. O receio de redução do papel do professor no processo de ensino aparece de forma heterogênea entre as áreas, porém sem evidência estatística robusta que permita sustentar diferenças sistemáticas associadas à grande área do conhecimento.

USO DA IA NAS DIFERENTES DIMENSÕES DOCENTES POR ÁREA DE CONHECIMENTO

Foram utilizados, novamente, os testes exato de Fisher (Freeman–Halton, 3×2) e do χ^2 ($\alpha = 0,05$; tabelas 3×2) para avaliar o uso da IA nas diferentes dimensões da atuação docente por áreas de conhecimento. Utilizou-se o agrupamento de áreas do conhecimento como anteriormente. Os testes ficaram, então, com 3 linhas: exatas, humanas e biológicas. As respostas sobre a utilização da IA foram agrupadas “utiliza mais” e “utiliza pouco ou não utiliza”. A hipótese testada para as diversas dimensões da atuação docente foi sempre a mesma: a de que a área de conhecimento do docente não influencia a maior ou menor utilização de IA.

Os resultados do teste de Fisher e de independência pelo *qui* são mostrados na Tabela 5. Há associação significativa em apenas um caso: existe associação significativa entre grande área do conhecimento e uso de IA em atividades de gestão universitária. Observa-se maior probabilidade de uso intensivo ou moderado entre docentes de humanas, menor probabilidade entre docentes de biológicas e padrão intermediário nas exatas. A associação detectada pelo teste de Fisher revela um gradiente entre áreas, com uso crescente da IA na gestão universitária das Ciências Biológicas para as Ciências Humanas.

Tabela 5. Uso nas diferentes dimensões: resultado dos testes.

Dimensão de atuação	Teste de Fischer		Teste do χ^2	
	p-valor	H ₀ é aceita?	p-valor	H ₀ é aceita?
Ensino (graduação/pós)	0,83	Sim	0,86	Sim
Pesquisa	0,27	Sim	0,31	Sim
Extensão universitária	0,71	Sim	0,75	Sim
Gestão universitária	0,04	Não	0,046	Não
Inovação, cultura, divulgação científica	0,19	Sim	0,23	Sim
Outras dimensões	0,52	Sim	0,056	Sim

O teste de independência pelo χ^2 corrobora o teste de Fischer, indicando associação estatisticamente significativa entre grande área do conhecimento e nível de uso da IA apenas na dimensão de gestão universitária.

Nas demais dimensões (ensino, pesquisa, extensão, inovação e outras) não se observou dependência estatística, sugerindo padrões de uso semelhantes entre as áreas, considerando o tamanho amostral do estudo.

CONSIDERAÇÕES FINAIS

O longo histórico da Unesp na formação docente, consolidado por iniciativas como o NEPP, CENEPP, IEP3 e, mais recentemente, pela CDeP3 e EFD, credencia a universidade a enfrentar o desafio atual da formação docente para o uso crítico e ético da Inteligência Artificial. Essa trajetória demonstra capacidade institucional para articular políticas, programas e ações que unem fundamentação pedagógica, atualização tecnológica e compromisso social.

Os resultados dos Ciclos de Oficinas Pedagógicas em Tecnologias Digitais evidenciam avanços relevantes na sensibilização e capacitação inicial dos docentes para uso da IA, mas apontam a necessidade de aprofundar e dar maior visibilidade às políticas institucionais já existentes a respeito do uso de IA, como a Resolução Unesp nº 13/2025 e a Portaria Unesp nº 125/2025. Fica claro que, embora haja concordância sobre a importância dessas diretrizes, parte significativa da comunidade docente não se sente plenamente amparada por elas. Isso reforça a urgência de ações de comunicação e formação que permitam compreender não apenas o que a política coloca como diretriz, mas como aplicá-la no cotidiano acadêmico.

Reafirma-se, portanto, a necessidade de manter e expandir ações de formação continuada em IA, contemplando todas as dimensões da atuação docente (ensino, pesquisa, extensão, gestão e inovação) e fortalecendo a criação de espaços de troca e colaboração. O protagonismo dos Núcleos Locais é determinante nesse processo: por estarem próximos à realidade e às especificidades de cada unidade, são estratégicos para contextualizar conteúdos, fomentar engajamento e garantir capilaridade às iniciativas.

Para alcançar o uso qualificado da IA na prática pedagógica, será essencial: i) ampliar o envolvimento dos Núcleos Locais como gestores e multiplicadores das ações formativas; ii) aumentar a visibilidade e compreensão das políticas institucionais de IA, relacionando princípios éticos e normativos à prática docente; iii) investir em engajamento e apoio ao corpo docente para superar receios identificados, como questões de autoria, plágio e confiabilidade das respostas geradas por IA;

e iv) assegurar a disponibilidade de ferramentas institucionais seguras e acessíveis que viabilizem a adoção da IA de forma ampla e ética.

Combinando a experiência acumulada e a estrutura institucional existente, a Unesp tem plenas condições de liderar, em âmbito nacional, a construção de uma cultura acadêmica que valoriza a formação profissional do seu corpo docente e que integra a IA como recurso educacional de forma responsável, crítica e inovadora.

REFERÊNCIAS

- CAMPOS, Vanessa T. Bueno; MORAIS, Sarah Juvencino de Oliveira. Desenvolvimento Profissional Docente na Educação Superior: constituição e contribuições de espaços formativos à docência universitária. *Ensino em Re-Vista*, v. 31, n. Contínua, p. 1–26, 2024. DOI: 10.14393/ER-v31e2024-19.
- DURSO, Samuel de Oliveira. Reflexões sobre a aplicação da inteligência artificial na educação e seus impactos para a atuação docente. *Educ. rev.*, Belo Horizonte, v. 40, e47980, 2024. DOI: 10.1590/0102-469847980.
- KUBOVICS, Michal. The impact of age groups' attitudes towards artificial intelligence and data protection. *International Journal of Business and Economics Research*, v. 14(3), 2025. DOI: 10.11648/j.ijber.20251403.13.
- MÉNDEZ-SUÁREZ, Mariano; MONFORT, Abel; HERVAS-OLIVER, Jose-Luis. Are you adopting artificial intelligence products? Social-demographic factors to explain customer acceptance. *European Research on Management and Business Economics*, v. 29(3), 100223, 2023. DOI: 10.1016/j.iedeen.2023.100223.
- SALVADOR, Maria Aparecida Tenório; SORDI, Mara Regina Lemes de. A formação docente para o ensino superior: impasses e perspectivas para a contínua aprendizagem. *Educar em Revista*, Curitiba, v. 40, e91566, 2024. DOI: 10.1590/1984-0411.91566
- UNESP (Universidade Estadual Paulista). *Resolução n° 45, de 03 de outubro de 2008*. Dispõe sobre a criação do Núcleo de Estudos e Práticas Pedagógicas da Unesp – NEPP/Unesp – junto da Pró-Reitoria de Graduação da Unesp. São Paulo, SP: Unesp, 2008. Disponível em <https://sistemas.unesp.br/legislacao-web/?base=R&numero=45&ano=2008&dataDocumento=03/10/2008>. Acesso em 16 dez. 2025.

UNESP (Universidade Estadual Paulista). *Resolução n° 74, de 22 de dezembro de 2022*. Cria a Coordenadoria de Desenvolvimento Profissional e Práticas Pedagógicas – Professora Adriana Chaves (CDeP3). São Paulo, SP: Unesp, 2022. Disponível em <https://sistemas.unesp.br/legislacao-web/?base=R&numero=74&ano=2022&dataDocumento=22/12/2022>. Acesso em 11 dez. 2025.

UNESP (Universidade Estadual Paulista). *Resolução Unesp n° 13, de 22 de abril de 2025*. Dispõe sobre a utilização da Inteligência Artificial na Universidade Estadual Paulista “Júlio de Mesquita Filho” (Unesp) e dá outras providências. São Paulo, SP: Unesp, 2025. Disponível em <https://sistemas.unesp.br/legislacao-web/?base=R&numero=13&ano=2025&dataDocumento=22/04/2025>. Acesso em 18 dez. 2025.

UNESP (Universidade Estadual Paulista). *Resolução Unesp n° 154, de 29 de novembro de 2012*. Dispõe sobre a criação do Centro de Estudos e Práticas Pedagógicas “Professora Adriana Josefa Ferreira Chaves” – CENEPP – Unesp. São Paulo, SP: Unesp, 2012. Disponível em <https://sistemas.unesp.br/legislacao-web/?base=R&numero=154&ano=2012&dataDocumento=29/11/2012>. Acesso em 5 jan. 2026.

UNESP (Universidade Estadual Paulista). *Resolução Unesp n° 23, de 17 de junho de 2025*. Dispõe sobre a criação da Escola de Formação dos Docentes da Unesp e institui o Programa de Formação Pedagógica e Desenvolvimento Profissional Docente (PROFOR). São Paulo, SP: Unesp, 2025. Disponível em <https://sistemas.unesp.br/legislacao-web/?base=R&numero=23&ano=2025&dataDocumento=17/06/2025>. Acesso em 18 dez. 2025.

UNESP (Universidade Estadual Paulista). *Resolução Unesp n° 79, de 05 de dezembro de 2018*. Dispõe sobre a transformação do Centro de Estudos e Práticas Pedagógicas “Professora Adriana Josefa Ferreira Chaves” – Cenepp, em “Instituto de Educação e Pesquisa em Práticas Pedagógica – IEP3 (Profa. Adriana Chaves)”, integrando à sua estrutura o Núcleo de Educação a Distância – NeaD. São Paulo, SP: Unesp, 2018. Disponível em <https://sistemas.unesp.br/legislacao-web/?base=R&numero=79&ano=2018&dataDocumento=05/12/2018>. Acesso em 5 jan. 2026.

VIEIRA, Géssika Mendes; VIEIRA, Vania Maria de Oliveira. Formação de Professores no Ensino Superior: a profissionalização docente por meio da pedagogia universitária. *Revista Leia Escola*, Campina Grande, v. 23, n. 2, p. 55–67, 2023. DOI: 10.5281/zenodo.10443747.

**ÉTICA NA INTELIGÊNCIA ARTIFICIAL: DESAFIOS
E RECOMENDAÇÕES - UMA PERSPECTIVA DE
DIREITOS HUMANOS E IGUALDADE DE GÊNERO**

**ETHICS IN ARTIFICIAL INTELLIGENCE:
CHALLENGES AND RECOMMENDATIONS - A HUMAN
RIGHTS AND GENDER EQUALITY PERSPECTIVE**

DOI: 10.29327/5798915.1-8

*Ligia Pavan Baptista*¹³

Resumo: A filosofia, sobretudo a filosofia prática, entendida como a ética e a filosofia política, tem um papel importante para a análise de questões contemporâneas decorrentes de novas tecnologias, tais como, a Inteligência Artificial. Se, por um lado, a IA representa uma significativa mudança de paradigma que certamente beneficia o desenvolvimento econômico e social em todos seus diversos aspectos e estes, por si só, são direitos humanos fundamentais, por outro, representa uma série de desafios complexos se considerarmos o tema do ponto de vista ético e vieses, tais como, riscos de desemprego, discriminação, racismo algorítmico, riscos à privacidade com o compartilhamento de dados e sistemas controversos como reconhecimento facial, desigualdade de acesso à nova tecnologia, desigualdade de gênero, desinformação, direitos autorais, ameaça à democracia, responsabilização de máquinas autônomas, dentre outros. Questões essas que estão demonstrando – por parte dos países, sejam de forma isoladas ou em blocos, como União Europeia e G20, organismos internacionais, como UNESCO, ONGs, grupos de pesquisa, universidades, redes de pesquisa nacionais e inter-

¹³ Professora de Ética e Filosofia Política e Programa de Pós-Graduação - Universidade de Brasília, ligia.pavanbaptista@gmail.com.

nacionais – a necessidade de regulamentação da Inteligência Artificial para que sua utilização seja responsável, socialmente justa e ambientalmente sustentável e, sobretudo, que seja ela centrada no ser humano, fundamentada nos direitos humanos e em linha com os Objetivos de Desenvolvimento Sustentável da Agenda 2030 das Nações Unidas.

Palavras-chave: Inteligência Artificial, Ética, Direitos Humanos, Tecnologia, Regulamentação.

Abstract: Philosophy, especially practical philosophy, understood as ethics and political philosophy, plays an important role in analyzing contemporary issues arising from new technologies, such as Artificial Intelligence. While AI represents a significant paradigm shift that certainly benefits economic and social development in all its diverse aspects—and these, in themselves, are fundamental human rights—it also presents a series of complex challenges when considering the topic from an ethical and biased perspective, such as risks of unemployment, discrimination, algorithmic racism, privacy risks from data sharing and controversial systems like facial recognition, unequal access to new technology, gender inequality, misinformation, copyright issues, threats to democracy, accountability of autonomous machines, among others. These issues are demonstrating—on the part of countries, whether individually or in blocs such as the European Union and the G20, international organizations such as UNESCO, NGOs, research groups, universities, national and international research networks—the need for regulation of Artificial Intelligence so that its use is responsible, socially just and environmentally sustainable and, above all, that it is human-centered, based on human rights and in line with the Sustainable Development Goals of the United Nations 2030 Agenda.

Keywords: Artificial Intelligence, Ethics, Human Rights, Technology, Regulation

INTRODUÇÃO

Na sequência do contexto histórico marcado pelo final da II Guerra Mundial, ainda sofrendo os impactos da primeira Guerra e as atrocidades que o mundo presenciou em termos de destruição, sofrimento e mortes A Organização das Nações Unidas, fundada em 1945, com a Carta de São Francisco, foi criada com o propósito de garantir a paz e a defesa, assim como evitar a possibilidade de ocorrência de um terceiro conflito mundial. A ideia central, baseada originalmente numa liga de Nações e, posteriormente, numa Federação de Estados livres, modelo apresentado por Kant em seu ensaio filosófico *À Paz Perpétua* – um projeto filosófico, já apresenta a busca da paz como condição imprescindível para a garantia dos direitos humanos.

No mesmo contexto histórico, três anos após a fundação das Nações Unidas, a Declaração Universal dos Direitos Humanos foi adotada e proclamada pela Assembleia Geral da ONU na Resolução 217-A (III) de 10 de dezembro de 1948, precedida historicamente pela *Déclaration des droit de l'homme et du citoyen*, votada pela Assembleia Nacional francesa, em 1789, sendo esta, por sua vez, precedida pelos *Bills of rights*, de 1776, das colônias americanas contra o domínio inglês e *Bill of rights* inglês, que consagrava a Revolução de 1680, sendo estes, por sua vez, influenciados pelo jusnaturalismo apresentado, principalmente, pelas teorias contratualistas de Hobbes, Locke e Rousseau.

Em referência à data em que a Declaração foi aprovada pela Assembleia Geral das Nações Unidas, o dia 10 de dezembro passou a ser celebrado como o Dia Internacional dos Direitos Humanos. O documento de trinta artigos, apontado por Bobbio como sendo o mais importante da história da humanidade, está celebrando, neste ano, 78 anos de existência. Ambos, a Carta das Nações Unidas e a Declaração Universal dos Direitos Humanos, podem ser vistos como documentos complementares que marcam, em conjunto, uma nova era do Direito Internacional, com referências recíprocas e princípios, respectivamente, endereçados aos Estados e aos indivíduos em geral, independente de suas nacionalida-

des ou quaisquer outras formas de distinção, tais como sexo, religião, opinião política, nacionalidade, condição social, língua ou outra distinção de qualquer espécie.

Partindo do princípio da universalidade, ou seja, que certos direitos devem ser reconhecidos universalmente, independente de qualquer circunstância ou fato que se possa considerar, a Declaração, pautada essencialmente na racionalidade como característica comum que define o gênero humano, reforça essencialmente o direito à vida, não somente do ponto de vista da mera sobrevivência, mas sim o direito de todos a uma vida digna, que implica, sobretudo, o direito à liberdade e à igualdade como direitos humanos universais e fundamentais, dos quais todos os demais direitos humanos se apresentam como condição de realização plena do indivíduo: o direito ao trabalho, à propriedade, o direito à ampla defesa e ao devido processo legal, o direito à nacionalidade, à educação, ao seguro desemprego, à saúde, à segurança, à liberdade de expressão e opinião, à propriedade e seu caráter inviolável, ao mesmo tempo em que condena suas violações, tais como, todas as formas de escravidão, ou qualquer restrição à liberdade, a prisão arbitrária, a desigualdade, a discriminação, o racismo, a violência, a tortura, o sofrimento, a censura e todas as demais violações que ameaçam ou impedem a plena realização da vida humana e as condições para seu desenvolvimento. Da mesma forma, ao definir no primeiro artigo que todos os seres humanos nascem livres e iguais em dignidade e direitos, é no princípio da racionalidade, fator comum e específico ao gênero humano, o elemento que fundamenta, no mesmo o princípio da fraternidade como regra geral de ação recíproca que deve guiar as relações humanas.

“Artigo I – Todos os seres humanos nascem livres e iguais em dignidade e direitos. São dotados de razão e consciência e devem agir em relação uns aos outros com espírito de fraternidade.” Da mesma forma que a paz entre os Estados, objetivo das Nações Unidas se mostra como a condição primeira e fundamental de garantia de realização dos direitos humanos, os princípios da Declaração, há 78 anos, é o fundamento de todo e qualquer Estado Democrático de Direito.

Desde 2015 os mesmos estão reforçados na agenda 2030 das Nações Unidas, com suas dezessete Metas de Desenvolvimento Sustentável, igualmente de caráter universal.

A IA pode ajudar a humanidade a enfrentar muitos problemas sociais graves, tais como a desigualdade social, e suas implicações e impactos no desenvolvimento de diversas áreas como educação, saúde, cultura, defesa, ciência, dentre outras, mas ao mesmo tempo pode representar uma série de desafios complexos em matéria de ética, direitos humanos, tais como, riscos de desemprego, discriminação, racismo algorítmico, riscos à privacidade com o compartilhamento de dados e sistemas controversos como reconhecimento facial, desigualdade de acesso à nova tecnologia, desinformação, dentre outros. A filosofia, sobretudo a filosofia prática, entendida como a ética e a filosofia política, tem um papel importante para a análise de questões contemporâneas decorrentes de novas tecnologias como a Inteligência Artificial.

Desde 2019, a UNESCO reforça papel da filosofia prática na análise do tema, tendo convidado a Rede de Mulheres Filósofas da América Latina, por ocasião dos 70 anos da presença da UNESCO na região, em cerimônia realizada em seu escritório em Montevidéu, a promover o painel *Ética na Inteligência Artificial, Promessas e Ameaças*, no XX Encontro da AFRA — Associação Filosófica da Argentina, realizado em *Mar del Plata*. Em 2021 foi aprovada, por 193 países membros das Nações Unidas, a *Recomendação sobre a Ética da Inteligência Artificial*, o primeiro instrumento global a definir normas para a IA, documento que foi traduzido para o português pela UNESCO Brasil em 2022. Em 2024 a UNESCO lançou o Observatório Global de Ética e Governança em IA. Dentre os inúmeros desafios, as recomendações expressam que IA deve ser usada e desenvolvida em benefício da humanidade no sentido de promover a justiça, a segurança, o acesso a todos, o uso de algoritmos de modo a reduzir e não a acentuar as desigualdades e as diversas formas de discriminação; e ainda tratar de questões relativas à privacidade de dados e à integridade da informação a igualdade de gênero e à democracia. Nossa pesquisa tem por ob-

jetivo desenvolver o tema da ética na Inteligência Artificial com base nas recomendações da UNESCO, questão que hoje domina o debate acadêmico internacional.

Dentre as redes nacionais e internacionais, sociedades científicas e grupos de pesquisa que relacionam a IA aos direitos humanos, sustentabilidade, igualdade de gênero e uso ético e responsável dos quais participamos destacamos:

- 1º. Women in IA Ethics – WAIE
- 2º. Women in IA
- 3º. Rede de Mulheres Filósofas da América Latina / UNESCO;
- 4º. AI for Education
- 5º. Center for AI and Digital Policy – CAIDP
- 6º. FORGES – Fórum de Gestão do Ensino Superior nos Países e Regiões de Língua Portuguesa
- 7º. SBPC – Sociedade Brasileira para o Progresso da Ciência
- 8ª. ANPOF – Associação Nacional de Pós-Graduação em Filosofia
- 9º. Grupo de Pesquisa : Direitos Humanos e Promoção da Igualdade de Gênero – ODS 5 / UnB
- 10º. UAI – Understanding Artificial Intelligence – Instituto de Estudos Avançados/USP
- 11º. Grupo de Pesquisa: Ideia: Inovação e Desafios Éticos da Inteligência Artificial – ITA/ CNPq

As três primeiras redes internacionais desenvolvendo pesquisas e eventos sobre o tema da IA na perspectiva de igualdade de gênero; a quarta, rede internacional dedicada a promover boas práticas de IA na Educação; a quinta, igualmente rede internacional que tem por objetivo promover políticas de boas práticas de IA e política digital e orientações para os países e blocos como G20; a sexta promove discussão sobre uso da IA na educação nos países e regiões da língua portuguesa tais como o I Colóquio Temático FORGES Inteligência Artificial e Tecnologias emergentes no Ensino Superior no Instituto de Educação da Universidade de Lisboa – Portugal (9 a 10 de fevereiro de 2026); a sétima, a SBPC, fundada em 1948, promove debate IA em suas publi-

cações, como o JC Notícias, e debates em suas reuniões anuais com especialistas de várias áreas do conhecimento permitindo um diálogo interdisciplinar como o tema exige, ainda pouco desenvolvido pelas instituições de educação superior no país; a oitava, ANPOF, fundada em 1983, promove encontros nacionais bianuais onde grupos de pesquisa na área de filosofia se encontram, dentre eles destacamos: GT História das Mulheres na Filosofia, GT Raça, Gênero e Classe e GT Filosofia e Gênero, onde o debate sobre IA, do ponto de vista dos direitos humanos e igualdade de gênero, está sendo introduzido a partir das questões éticas. O nono, Grupo de Pesquisa Direitos Humanos e Promoção da Igualdade de Gênero ODS 5 – grupo de pesquisa fundado na Universidade de Brasília em 2024, que tem como objetivo promover os direitos humanos e a ODS 5 – Promoção da Igualdade de Gênero, que produz o Boletim Chauí, atualmente em sua quinta edição; o décimo e o décimo primeiro vinculados a centros de excelência brasileiros, Instituto de Estudos Avançados da Universidade de São Paulo IEA/USP e Instituto Tecnológico da Aeronáutica (ITA).

O Instituto de Estudos Avançados/USP, além da rede UAI, que promove debates com especialistas de forma virtual, presencial e híbrida, além de publicações como a recém lançada obra coordenada pelo Prof. Glauco Arbix: *IA Responsável na Encruzilhada – Desafios Geopolíticos, Regulatórios e de Governança*. Destacamos que IEA/USP possui igualmente duas cátedras dedicadas à temática de IA: Cátedra Oscar Sala (IEA e CGI) e a recém criada Cátedra IA Responsável, em parceria com Google. O Instituto Tecnológico da Aeronáutica, além de ofertar pela primeira vez em 2024 a disciplina *Ética na Inteligência Artificial*, ministrada pelo Prof. Delmo Mattos da Silva, realizou no ano de 2025 *workshop* e série de seminários promovidos pelo Grupo de Pesquisa Ideia – Inovação e Desafios Éticos na IA ITA/CNPq, criado em 2025, que tem, como objetivo principal: compreender e problematizar os desafios éticos decorrentes do desenvolvimento e implementação de sistemas de IA relativos à transferência de tecnologia para inovação em IA.

São redes que buscam examinar os impactos éticos da IA em aspectos como privacidade, discriminação, justiça social e autonomia, visando garantir que o desenvolvimento e uso dessa tecnologia respeitem os direitos humanos, considerem as implicações morais de suas decisões e ações e promovam o bem-estar geral. Para tanto, objetiva-se identificar e avaliar os riscos éticos associados ao uso de IA, sobretudo aqueles associados ao viés algorítmico, privacidade, desigualdade econômica, segurança e confiabilidade mediante o debate multidisciplinar dos dilemas éticos decorrentes da sua identificação.

DESAFIOS ÉTICOS NA UTILIZAÇÃO DA IA: O ESTADO DA ARTE

Não há dúvida que a IA está alterando nossas vidas e é inquestionável que seu desenvolvimento pode trazer benefícios para que a sociedade possa enfrentar muitos de seus problemas sociais mais graves, tais como a desigualdade social, a pobreza, a miséria e a fome, além das mudanças climáticas e dos grandes desafios atuais. É indiscutível que suas implicações e impactos podem contribuir significativamente no desenvolvimento de diversas áreas, tais como, educação, saúde, cultura, arte, defesa, ciência dentre outras.

Ao mesmo tempo, pode representar uma série de desafios complexos em matéria de ética, sobretudo relacionada à garantia dos direitos humanos e liberdades fundamentais, no sentido de evitar suas violações, tais como, desemprego, discriminação, racismo algorítmico, riscos à privacidade com o compartilhamento de dados e sistemas controversos como reconhecimento facial, desigualdade de acesso à nova tecnologia, desinformação, desigualdade de gênero, dentre outros que estão demandando atenção crescente dos especialistas na área.

Ainda que o tema, por si só, gere controvérsias e debates, é possível observar que há consenso entre os especialistas sobre alguns aspectos relevantes da IA, dentre eles:

1. Deve ser desenvolvida com um único objetivo: a melhoria da qualidade de vida das pessoas e que seu desenvolvimento deve ser centrado no ser humano;

2. Seu desenvolvimento tem causado preocupação na medida em que envolve riscos, sobretudo relacionados aos direitos humanos e liberdades fundamentais;
3. A IA precisa ser regulamentada pelos países, essa regulamentação precisa encontrar um ponto de equilíbrio para evitar os riscos e, ao mesmo tempo, permitir a liberdade necessária ao seu desenvolvimento;
4. Muitos países, inclusive o Brasil, estão debatendo atualmente sua regulamentação;
5. As recomendações da UNESCO sobre ética em IA e seus princípios balizadores estão sendo utilizados nos debates sobre regulamentação pelos países que estão discutindo a questão em seus respectivos parlamentos;
6. A IA está se desenvolvendo muito mais rapidamente do que nossa capacidade de entender os desafios que representa e dar respostas a eles, por exemplo, por meio de instrumentos como regulamentações, recomendações, manual de boas práticas, orientações, diretrizes, códigos de ética, dentre outras práticas semelhantes.

Nesse sentido, a filosofia – sobretudo a filosofia prática, definida por Aristóteles como a área da filosofia que depende da ação humana, nas quais se inserem a ética e a filosofia política, entendidas pelo Filósofo respectivamente como o estudo da ação humana tendo em vista o bem privado, e o estudo da ação humana tendo em vista o bem público – tem um papel importante para a análise de questões contemporâneas decorrentes de novas tecnologias como a Inteligência Artificial que, se sabe, está já alterando de maneira extremamente veloz nossa realidade, nossas relações, nossa visão de mundo e até mesmo o que até então entendemos por humanidade, a partir da nossa concepção do que é um ser humano e de sua capacidade de criação. A UNESCO definiu o Dia Mundial da Filosofia desde 2005, celebrado na terceira quinta-feira de novembro por seus parceiros para enfatizar a importância dessa área do conhecimento tão antiga quanto a própria humanidade.

A Recomendação sobre a Ética da Inteligência Artificial da UNESCO é um instrumento global que define nor-

mas para o uso ético da IA. A recomendação foi aprovada em 2021 pela Conferência Geral da UNESCO e tem como objetivo garantir que a IA beneficie a humanidade e não a prejudique. A recomendação da UNESCO estabelece os seguintes princípios para o desenvolvimento e uso da IA:

1. Garantir a proteção e promoção dos direitos humanos e da dignidade humana;
2. Assegurar a diversidade e inclusão;
3. Evitar que os preconceitos do mundo real sejam reproduzidos no mundo virtual;
4. Promover a sustentabilidade ambiental;
5. Garantir a equidade;
6. Garantir que as decisões que afetam milhões de pessoas sejam justas, transparentes e contestáveis;
7. Garantir que os desenvolvimentos de IA obedeçam ao estado de direito;
8. A recomendação da UNESCO também estabelece que os sistemas de IA devem ser monitorados e avaliados de forma rigorosa e recomenda que os governos desenvolvam currículos de ética em IA para todos os níveis educacionais;

Destacamos que, em 2024, A UNESCO criou o Observatório Global de Ética e Governança em IA, lançado no Fórum Global sobre Ética da Inteligência Artificial.

Como instrumento regional destacamos a Lei de IA da EU – União Europeia, aprovada em maio de 2024. A lei é considerada a primeira do mundo a regular o uso de IA e estabelece um padrão global para a regulamentação da tecnologia. A Lei de IA da UE tem como objetivos:

1. Promover o desenvolvimento e a adoção de sistemas de IA seguros e confiáveis;
2. Priorizar o respeito aos direitos fundamentais dos cidadãos da UE;
3. Estimular investimentos e inovações em IA dentro da Europa;
4. A lei adota uma abordagem “baseada em riscos”, o que significa que sistemas de IA com maior risco para a sociedade terão regulamentações mais rigorosas;

5. A Lei de IA da UE entrou em vigor 20 dias após a sua publicação oficial no jornal oficial europeu, em 12 de julho de 2024; Nossa pesquisa tem por objetivo desenvolver o tema da ética na Inteligência Artificial – questão que hoje domina o interesse da mídia, o debate acadêmico e político nacional e internacional – com base na Recomendação da UNESCO, único instrumento global sobre o tema, aprovado em 2021 pelos 193 países membros da organização que, como tal, se comprometem a implementar tais recomendações.

O TEMA GERA OS SEGUINTE DESAFIOS:

1. Como conciliar o desenvolvimento de IA e a proteção dos direitos humanos e liberdades fundamentais?
2. Como conciliar o desenvolvimento da IA com os 17 Objetivos de Desenvolvimento Sustentável – Agenda 2030?
3. Como conciliar o desenvolvimento de IA com o ODS 5 – Promoção da Igualdade de Gênero, sem a qual não poderá ser considerada inclusiva?
4. Como conciliar o desenvolvimento de IA com o combate à desinformação, considerada hoje o maior risco à democracia?
5. Como conciliar o desenvolvimento de IA com a desigualdade e a exclusão no acesso a novas tecnologias?
6. IA nas Universidades – Desafios do uso de IA na Educação Superior: será preciso criar um Código de Ética de utilização IA nas Universidades? – Grupos de Pesquisa/CNPq.
7. IA no Brasil – Regulamentação /Projetos de Lei – Plano IA para o Bem de Todos.
8. Como conciliar o desenvolvimento de IA e a proteção dos direitos humanos e liberdades fundamentais?
9. Como conciliar o desenvolvimento da IA com os 17 Objetivos de Desenvolvimento Sustentável – Agenda 2030?

O ranking britânico THE – Times Higher Education lançou em 2022 o *Times Higher Education Impact Ranking*, que utiliza métricas de desempenho global que avaliam as universidades em relação aos Objetivos de Desenvolvimento Sustentável (ODS) das Nações Unidas, para avaliar as universidades especificamente por meio de medidas de implementação

dos 17 ODS. Como é possível conciliar o desenvolvimento de IA com o princípio ético da sustentabilidade? Especialistas já estudam o alto impacto do uso de IA no consumo de energia elétrica e água. Como o Brasil, reconhecido pela sua capacidade de energias renováveis, segundo especialistas ainda subutilizadas, pode contribuir para minimizar esse impacto? Podemos criar um selo: IA VERDE ou IA SUSTENTÁVEL para empresas públicas e privadas que de fato comprovem, por meios de ações efetivas, o compromisso com a sustentabilidade na utilização da IA

10. Como é possível conciliar o desenvolvimento de IA com o ODS 5 – Promoção da Igualdade de Gênero?

Esse foi o tema do encontro que participamos realizado na sede da UNESCO, em Paris, com transmissão ao vivo no dia 30 de outubro 2024.

11. Como conciliar o desenvolvimento de IA com o combate à desinformação e às *deepfakes*, consideradas hoje o maior risco à democracia?

Unesco e União Europeia não utilizam o termo “*Fake News*”, utilizam o termo “Desinformação”. O termo “*deepfake*” é utilizado como uma técnica que permite alterar um vídeo, foto, áudio com ajuda de inteligência artificial (IA).

Como a IA pode combater a desinformação e as *deepfakes* e promover a integridade da informação? Dentre os riscos associados à IA, como o impacto no mercado de trabalho; o uso militar; a massificação da desinformação; e a violação de privacidade, a disseminação de desinformação e das *deepfakes* representam uma ameaça significativa à democracia. Como conciliar o desenvolvimento de IA com o desafio de combater a desigualdade e a exclusão no acesso a novas tecnologias? Como IA pode ser uma tecnologia inclusiva?

IA NAS UNIVERSIDADES: DESAFIOS DO USO DE IA NA EDUCAÇÃO SUPERIOR

A Cátedra UNESCO de Ética na Inteligência Artificial e Governança foi fundada em 28 de outubro de 2024 na *IE University School of Humanities* – em Madrid na Espa-

nha – Coordenada pelo Professor Ted Lechterman, com a missão de promover a Recomendação sobre Ética na Inteligência Artificial da UNESCO.

O Primeiro curso de bacharelado em Inteligência Artificial no Brasil foi criado pelo Instituto de Informática da Universidade Federal de Goiás em 2019 – a primeira turma iniciou o curso em 2020. O curso tem o objetivo de formar profissionais aptos a resolver problemas complexos por meio da Inteligência Artificial, incluindo sistemas embarcados e autônomos, de forma inovadora e empreendedora. 85% dos estudantes têm bolsas de estudos pagas pelas empresas parceiras. A primeira turma de formandos, composta de 14 estudantes, concluiu o curso em 22 de março de 2024. Heloisy Pereira Rodrigues, de 24 anos, de Ceres/Goiás, é a primeira mulher formada no primeiro curso de graduação de Inteligência Artificial da UFG no Brasil.

IA foi tema da redação vestibular UnB – Universidade de Brasília, em 2024, estimulando o debate sobre uma questão polêmica que considera a tecnologia uma ameaça ao pensamento crítico, segundo Noah Chomshy, a partir de trecho de uma entrevista concedida por ele ao jornal português Público:

A IA, tal como é entendida atualmente, é um projeto corporativo que visa reunir conteúdos para serem usados por sistemas de simulação em grande escala. De que vale compreender o que quer que seja quando se pode analisar um sem-fim de dados e prever o que vai acontecer? Este é o mais radical ataque ao pensamento crítico, à inteligência crítica e particularmente à ciência que eu alguma vez vi (Fonte: Correio Braziliense, 24.11.2024).

IA NAS UNIVERSIDADES BRASILEIRAS

Universidades discutem regras de uso de IA e definem orientações. Será preciso criar um código de ética para definir uso adequado de IA nas universidades? A FAPESP sugere diretrizes para uso responsável de IA nas universidades:

1. TRANSPARÊNCIA

A utilização de ferramentas de IA, caso seja permitida, deve ser informada na metodologia do trabalho /artigos acadêmicos;

2. PROTEÇÃO DE DADOS

É preciso avaliar quais dados podem ser utilizados a partir do uso ferramentas de IA; Informações sensíveis ou inéditas não devem ser compartilhadas;

3. AUTORIA

A IA generativa não pode ser considerada autora de trabalhos acadêmicos e/ou artigos científicos. A autoria é sempre um ser humano que pode utilizar ferramentas de IA;

4. CLAREZA

As ementas das disciplinas devem informar com clareza o que os estudantes podem ou não fazer com a utilização de ferramentas de IA;

5. VIESES

As ferramentas de IA podem gerar desinformação, preconceitos, discriminação, racismo e é preciso avaliar com atenção os dados que são fornecidos por ferramentas de IA para evitar tais riscos.

Será preciso exigir a assinatura de uma declaração de pesquisadoras/pesquisadoras que se comprometem a utilizar IA com ética, responsabilidade e respeito aos direitos humanos, assim como já é o procedimento em relação à ética na pesquisa científica?

O ranking britânico THE – Times Higher Education lançou, em 2022, o *Times Higher Education Impact Ranking*, que utiliza métricas de desempenho global que avaliam as universidades em relação à promoção dos Objetivos de Desenvolvimento Sustentável (ODS) das Nações Unidas, avaliando, especificamente, medidas de implementação dos 17 ODS. Em 2024 analisou 2.152 universidades de 125 países/regiões.

As TOP/5 universidades mais bem avaliadas do mundo são:

1. Western Sydney University – Austrália
2. University of Manchester – Reino Unido
3. University of Tasmania – Austrália
4. Aalborg University – Dinamarca
5. RMIT University – Austrália

As TOP/5 universidades mais bem avaliadas no Brasil:

1. Universidade de São Paulo / USP – SP
2. Universidade Federal de Itajubá – MG
3. Universidade Estadual Paulista / UNESP – SP
4. Universidade Federal do Paraná/ UFPR – PR
5. Universidade de Campinas / UNICAMP – SP

GRUPOS DE PESQUISA EM IA NO BRASIL: DIRETÓRIO DE GRUPOS DE PESQUISA DO CNPQ

Busca feita somente pelo termo “Inteligência Artificial” no nome do grupo no Diretório de Grupos de Pesquisa do CNPq apresentou como resultado 125 Grupos de Pesquisa em Instituições de todo o país nas seguintes áreas do conhecimento: ciência da computação (45%); engenharias; direito; filosofia; educação; medicina e ciência da informação.

REGULAMENTAÇÃO DA IA NO BRASIL

No âmbito da regulamentação da IA no Brasil destacamos projetos de lei no Senado Federal e o Plano IA para o Bem de Todos.

Senado Federal criou em 2023 a Comissão Temporária de Inteligência Artificial no Brasil (CTIA) para tratar dos projetos de lei que tramitam em conjunto e tratam da regulamentação no Brasil.

PLs 5.051/2019 872/2021 e 2010/2024 dos princípios para uso de IA.

PL 5.691/2019 da política nacional dos fundamentos e diretrizes para o desenvolvimento e aplicação.

PL 21/2020 dos fundamentos e diretrizes para o desenvolvimento e aplicação.

PL 3.592/2023 de imagens e áudios de pessoas falecidas.

PL 145/2024 da publicidade enganosa.

PL 146/2024 do crime de falsa identidade.

PL 266/2024 da atuação de médicos, advogados e juízes.

Os PLs se fundamentam na regulamentação de IA da UNESCO e da União Europeia.

O Projeto de Lei 2338/2023 Normas para uso de IA foi lido pelo relator dia 28 de novembro de 2024 e aprovado dia 10 de dezembro de 2024.

PLANO IA PARA O BEM DE TODOS

O plano “IA para o bem de todos” foi elaborado por mais de cem instituições, dentre elas a Academia Brasileira de Ciências (ABC) e a Sociedade Brasileira para o Progresso da Ciência (SBPC). O documento foi entregue ao presidente Lula na cerimônia de Abertura da 5ª Conferência Nacional de Ciência, Tecnologia e Inovação, realizada pelo Ministério da Ciência, Tecnologia e Inovação – MCTI, em Brasília, em 30 de julho 2024. O evento, que ocorreu 30 de julho a 3 de agosto 2024, apresentou painéis sobre IA. A sistematização dos painéis do evento foi feita por IA, com revisão humana. O Plano tem por objetivo impulsionar o desenvolvimento tecnológico e a aplicação de IA no Brasil em benefício e toda a sociedade, promovendo inovações sustentáveis e inclusivas, gerando oportunidades de emprego e melhorando a qualidade de vida dos brasileiros.

O PLANO SE ESTRUTURA EM 5 EIXOS

1. Infraestrutura e Desenvolvimento de IA;
2. Difusão, Formação e Capacitação em IA;
3. IA para Melhoria dos Serviços Públicos;
4. IA para Inovação Empresarial;
5. Apoio ao Processo Regulatório e de Governança de IA.

O PLANO SE ESTRUTURA EM 6 PRINCÍPIOS FUNDAMENTAIS

1. IA centrada no ser humano;
2. IA para o bem de todos;
3. IA ética e responsável;

4. IA sustentável;
5. IA para o desenvolvimento nacional;
6. IA democrática e participativa.

O Plano Brasileiro de “IA para o bem de todos” terá um supercomputador e investimentos de R\$ 23 bilhões no período de 4 anos: 2024– 2028.

EIXO 2 – ÁREA DE ATUAÇÃO DAS INSTITUIÇÕES DE ENSINO SUPERIOR

DIFUSÃO, FORMAÇÃO E CAPACITAÇÃO EM IA

Recurso de R\$ 1,5 bilhão para campanhas de popularização como Olimpíada Brasileira de IA, formação em todos os níveis, criação de disciplinas, cursos de graduação, pós-graduação, especialização, capacitação com oferta de bolsas de estudo e para pesquisa em IA no Brasil e no exterior, qualificação profissional, cursos virtuais e presenciais em IA, realização de eventos científicos e projetos de letramento e inclusão digital. Podemos concluir que não há ainda um consenso por parte dos especialistas sobre determinados aspectos da IA, dentre eles destacamos:

1. Se IA é uma tecnologia ou uma ciência;
2. Se a Inteligência Artificial é um inteligente e, tampouco, se é artificial;
3. Se é uma nova tecnologia ou se é uma tecnologia antiga (pode-se dizer que já tem mais de 50 anos);
4. Se é uma ferramenta (para aqueles que consideram que uma máquina não é capaz de pensar) ou um agente (para aqueles que consideram que é possível uma máquina pensar ou pelo menos decidir);
5. Que IA não sendo considerada um agente moral (não tem noção de certo/errado – justo/injusto), fica em aberto a questão da responsabilização de uma máquina autônoma do ponto de vista jurídico.

Por outro lado, podemos afirmar, que há um certo consenso a respeito sobre os seguintes pontos:

1. Que IA é uma área de conhecimento interdisciplinar/multidisciplinar/transdisciplinar e que por essa razão exige uma maior interação entre diversas áreas do conhecimento;
2. Que há uma crescente demanda no mercado de trabalho para especialistas em IA;
3. Que seu desenvolvimento envolve riscos e que é preciso criar mecanismos para que sejam minimizados;
4. Que IA precisa ser regulamentada; que cada país é soberano para decidir sobre essa questão; e que a soberania reside no direito de escolha sobre a necessidade de criar ou não uma regulamentação própria e de definir seu conteúdo, no caso de ser escolhida a primeira opção;
5. Que por seu caráter interdisciplinar as ciências humanas e sociais estão sendo demandadas a abordarem o tema em conjunto com outras áreas de conhecimento;
6. Que IA representa uma nova área de pesquisa para a filosofia, sobretudo, a filosofia prática: a ética e a filosofia política. Especificamente, a pesquisa sobre IA em filosofia se encaixa na área de “Ética Aplicada” ou “Ética Prática” ou ainda “Ética e Tecnologia”, ainda pouco desenvolvida nos cursos de filosofia no Brasil, e estamos sendo demandados a desenvolver pesquisas e a dar respostas para os novos desafios que a IA apresenta.

Especialmente, é consenso que, para ser considerada do ponto de vista da ética, a IA deve ser desenvolvida em qualquer área que se aplique, tanto na esfera privada quanto pública, com os seguintes objetivos: para o bem de todos; de forma sustentável, socialmente justa, inclusiva, transparente e responsável; simultaneamente com políticas públicas de inclusão digital e promoção da igualdade de gênero; e em linha com os 17 Objetivos de Desenvolvimento Sustentável. Finalmente, pode-se concluir que é consenso geral que, em seu desenvolvimento, a IA deve ser centrada no ser humano, ser utilizada de forma responsável e ética e ter como base e condição o respeito aos direitos humanos e liberdades fundamentais.

Para concluir, ressaltamos que, a partir dos recursos disponibilizados pelo Plano Brasileiro IA para o Bem de To-

dos, as universidades brasileiras estão investindo de forma estruturada em áreas estratégicas para o desenvolvimento científico, econômico e social no país, investindo em supercomputadores, melhoria da conectividade em seus campi e criação de cursos de bacharelado de caráter interdisciplinar em Inteligência Artificial, dentre elas, UFF – Universidade Federal Fluminense, UnB – Universidade de Brasília, UFRN – Universidade Federal do Rio Grande do Norte. Reitor da USP recentemente empossado, Aluísio Segurado, informa que será criado núcleo de transformação digital e disciplina de IA em todos os cursos de graduação da instituição. CNE – Conselho Nacional de Educação informa que aprovou diretriz para que IA seja introduzida nos currículos das escolas de ensino médio em todo o país.

Finalmente, ressaltamos que o Brasil está em lugar de destaque internacional com o prêmio UNESCO-Uzbequistão para pesquisa científica sobre ética na Inteligência Artificial, recebido pelo Prof. Virgílio Almeida (UFMG – IEA/USP) em 7 de novembro de 2025, por sua contribuição para políticas globais para o uso responsável da Inteligência Artificial.

REFERÊNCIAS

ALMEIDA, Virgílio; MENDONÇA, Ricardo; FILGUEIRAS, Fernando. *Política dos algoritmos: instituições e as transformações da vida social*. São Paulo: UBU, 2025.

BRASIL. MCTI. *Plano IA para o Bem de Todos*. Disponível em: https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/cct/legislacao/arquivos/IA_para_o_Bem_de_Todos.pdf. Acesso em 26 jan. 2026.

CNPq. *Diretório de Grupos de Pesquisa*. Disponível em: <https://lattes.cnpq.br/web/dgp>. Acesso em: 17 dez. 2025

COECKELBERGH, Mark. *Ética na Inteligência Artificial*. São Paulo: UBU, 2024.

KAUFMAN, Dora. *Desmistificando a Inteligência Artificial*. Belo Horizonte: Autêntica, 2025.

UNICEF. *Declaração Universal dos Direitos Humanos*, 1948. Disponível em: <https://www.unicef.org/brazil/declaracao-universal-dos-direitos-humanos>. Acesso em: 17 dez. 2025.

**DO COLONIALISMO À SOBERANIA DIGITAL: ÉTICA
E JUSTIÇA PARA O FLORESCIMENTO HUMANO**

**FROM COLONIALISM TO DIGITAL SOVEREIGNTY:
ETHICS AND JUSTICE FOR HUMAN FLOURISHING**

DOI: 10.29327/5798915.1-6

*Rafael Evangelista*¹⁴

*Rodolfo Avelino*¹⁵

Resumo: Este texto parte de um deslocamento da ideia de ética, como frequentemente tem sido empregada nas discussões sobre tecnologia, para sustentar uma discussão sobre como as tecnologias da informação (o digital) podem ser empregadas de maneira a permitirem o florescimento humano. Para isso, adota um posicionamento situado a partir do Sul Global, colocando o colonialismo e a colonialidade digitais como um dos obstáculos principais para as sociedades minorizadas. Postula então os debates e a reivindicação por “soberania digital” como reações diretas ao reconhecimento acadêmico do colonialismo digital, formulando a ideia de soberania digital como em continuidade com uma perspectiva de florescimento humano.

Palavras-chave: Inteligência Artificial; Florescimento Humano; Sul Global; Colonialismo Digital; Soberania Digital

Abstract: This text starts from a shift in the idea of ethics, as it has often been employed in discussions about technology, to sustain a discussion on how information technologies (the digital) can be used in ways that allow for human flourishing. To this end, it adopts a situated position from the Global South,

¹⁴ Professor do Programa de Pós-Graduação em Divulgação Científica e Cultural da Unicamp, rae@unicamp.br.

¹⁵ Professor no curso de Engenharia e da Ciência da Computação do Insper, rodolfosal@insper.edu.br.

identifying digital colonialism and coloniality as one of the main obstacles faced by minoritized societies. Then it posits debates and claims around “digital sovereignty” as direct reactions to the academic recognition of digital colonialism, formulating the idea of digital sovereignty as being in continuity with a perspective of human flourishing.

Keywords: Artificial Intelligence; Human Flourishing; Global South; Digital Colonialism; Digital Sovereignty

INTRODUÇÃO

Os debates sobre os efeitos sociais, em geral negativos, da Inteligência Artificial (IA) têm sido invariavelmente acompanhados por uma discussão sobre “ética” ou “uso ético”. A palavra ecoa uma preocupação pertinente de diversos setores da sociedade sobre como aproveitar socialmente eventuais benefícios dos sistemas de IA sem que os malefícios fujam do controle. Ao mesmo tempo, a palavra ética é sacada, em especial por setores empresariais, como meio para fazer avançar a ideia de que uma espécie de “bom comportamento” ou comportamento ético dos desenvolvedores e implementadores de tecnologias, que de alguma forma tornariam desnecessárias, por parte do Estado, medidas regulatórias, restritivas ou punitivas que de fato disciplinassem e controlassem o desenvolvimento e a aplicação da IA. A partir de entrevistas com empresários da Europa central, Baker-Brunnbauer (2021) aponta uma considerável dispersão de sentidos entre eles sobre as definições de ética e moral. Uma constante, no entanto, foi percepção limitada sobre ética, basicamente ligada à imagem externa da empresa relacionada à responsabilidade social. Apenas danos diretos e imediatos a consumidores foram tomados como dignos de consideração.

Um debate analítico, significativo e propositivo sobre as consequências possivelmente transformadoras das tecnologias precisa articular cenários conjunturais, conceitos claros e um posicionamento localizado objetivo (Haraway, 1995). É nesse sentido que inicialmente propomos, respectivamente, uma discussão sobre colonialismo e colonialidade digital; justiça de dados e florescimento humano; e os limites das demandas por ética se a perspectiva que se adota é a das populações minorizadas.

ALINHAMENTOS CONCEITUAIS

Ainda que o termo mais frequentemente mobilizado nos debates públicos seja IA, o que está em jogo é muito mais amplo do que os sistemas generativos recentes, como o ChatGPT. O conceito de inteligência artificial, entendido como a capacidade de sistemas computacionais produzirem classificações, ordenações, inferências e conteúdos a partir de grandes volumes de dados, antecede em muito a onda atual de modelos generativos. Como argumenta Crawford (2021), a IA deve ser compreendida não como uma tecnologia isolada, mas como um sistema sociotécnico que envolve infraestruturas, cadeias de extração, trabalho humano, práticas institucionais e relações de poder que moldam tanto sua produção quanto seus efeitos. Na verdade, os arcabouços que sustentam essa forma de análise e decisão automatizada – sobretudo o aprendizado de máquina, a mineração de dados, os sistemas de recomendação e a modelagem preditiva – tornaram-se parte central da vida social ao menos desde meados dos anos 2000, com a consolidação das infraestruturas de *big data* e a consolidação do capitalismo de vigilância (Zuboff, 2015) como lógica basilar das grandes plataformas. Assim, a emergência dos modelos generativos é apenas a manifestação mais visível de um processo mais longo de intensificação da extração e processamento de dados, que reorganizam modos de conhecer, governar e interagir.

Nesse sentido, o debate sobre colonialismo e colonialidade digital é indissociável da discussão sobre IA. Qualquer discussão sobre ética precisa levar em conta os efeitos assimétricos que a digitalização vem produzindo em diferentes grupos populacionais.

Antonio Casili (2017) nos ajuda a inventariar as primeiras menções a colonialismo nos debates sobre tecnologia digital. Inicialmente o termo colonialismo passou a ser mobilizado sobretudo como metáfora crítica, procurando capturar a expansão de formas de governo capitalista próprias da economia de plataformas. Um dos usos iniciais é o de Casati (2013 *apud* Casilli, 2017), para quem o “colonialismo digital” nomeia a normatividade automática instaurada pelas em-

presas de tecnologia – a crença, fundada em determinismo tecnológico, de que toda a vida social deve ser mediada por sistemas conectados. Nesse enquadramento, o colonialismo refere-se a políticas agressivas e decisões econômicas discricionárias que moldam comportamentos e instituem assimetrias de poder, ainda que Casati não explore em detalhe como esses processos se enraízam em modelos de negócio, cadeias de valor e regimes de trabalho intensivo em dados. Nesse sentido, colonialismo está bastante próximo do que, mais tarde, van Dijck (2014) já chamava de dadoísmo (*dataism*), uma crença generalizada na quantificação e sua objetividade, somada a uma confiança no potencial do rastreamento de todos os tipos de comportamento humano e social através de tecnologias informacionais *on-line*. Ela aponta ainda a confiança nos agentes institucionais em torno dessas operações, tanto nos que coletam, interpretam e compartilham dados como aqueles que deveriam zelar para que essas atividades tivessem uso socialmente justo. Outro uso inicial mapeado por Casati é o de Kleiner (2016, apud Casilli, 2017), que descreve a passagem de uma internet originalmente descentralizada e comunitária para uma rede centralizada e cercada por entidades corporativas oligopolísticas. Aqui, a colonização digital designa a apropriação predatória das infraestruturas de comunicação, que reintermedeiam e homogenizam não apenas mercadorias, mas também as normas e práticas sociais que lhes dão contexto.

Como argumenta Nyabola (2023), a colonialidade digital não se expressa apenas na extração de dados ou no controle de infraestruturas, mas de forma particularmente aguda nas fronteiras, onde tecnologias biométricas, IA e vigilância criam verdadeiras “culturas de fronteira” carcerárias. É nesse espaço que as continuidades coloniais se atualizam com mais violência, projetando poder sobre populações historicamente vulnerabilizadas.

O diagnóstico de Nyabola sobre a fronteira como um espaço privilegiado de exercício de violência digital se fortalece quando conectado à leitura estrutural de Kwet (2024). Para ele, o colonialismo digital opera por meio do controle

das infraestruturas globais – cabos submarinos, data centers, sistemas operacionais, nuvens corporativas – que permitem a projeção extraterritorial de poder sobre outros Estados. Assim, o que Nyabola observa nas fronteiras, Kwet identifica no plano sistêmico: a capacidade dos EUA e das *Big Techs* de definirem as condições materiais que tornam possíveis tecnologias de vigilância e controle migratório.

No contexto acadêmico latino-americano, ainda que já houvesse discussões relevantes sobre as assimetrias de poder entre Norte e Sul no campo digital – e, mais especificamente, sobre as tecnologias de vigilância – desde meados de 2015, mas com maior visibilidade a partir de 2019 nos simpósios da Rede Latino-Americana de Estudos em Vigilância, Tecnologia e Sociedade (Lavits), foi o texto de Couldry e Mejias (2019), *Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject*, que acabou por concentrar parte significativa do debate. Os autores formulam o conceito de colonialismo de dados, tomando como ponto de partida a colonização histórica – centrada na extração de recursos naturais e humanos e na constituição das bases do capitalismo industrial – para argumentar que a atual economia de dados institui um novo regime de apropriação: não mais de terras ou corpos, mas da própria vida cotidiana, formulada como relações de dados implementadas pelas grandes plataformas, fazendo da vida cotidiana matéria-prima informacional. No entanto, embora reconheçam a persistência de práticas exploratórias que marcam a relação estrutural entre Norte e Sul, sua análise tende a deslocar essa assimetria para um plano secundário. Ao ampliar o escopo do colonialismo de dados para englobar dinâmicas que atravessam também as populações do Norte Global, o conceito perde precisão para descrever as desigualdades históricas e materiais que moldam, de forma diferenciada, as condições de extração, vulnerabilidade e dependência no Sul Global.

No mesmo ano, 2019, o pesquisador sul-africano Michael Kwet emprega o termo colonialismo digital em um texto de influência crescente na academia e na sociedade civil brasileiras, especialmente por explicitar as profundas as-

simetrias políticas que estruturam o ambiente digital contemporâneo. Diferentemente de Couldry e Mejias, cujo foco recai sobretudo sobre as “relações de dados”, Kwet enfatiza que o cerne do problema está no controle das infraestruturas – e não apenas na produção ou circulação dos dados. Para ele, as formas contemporâneas de dominação digital se inscrevem em uma continuidade direta com o sistema colonial: trata-se de um poder centralizado que opera por meio do domínio de três pilares fundamentais do ecossistema digital – o *software* (controle do código, do desenvolvimento e da propriedade intelectual), o *hardware* (controle de servidores e centros de dados) e a conectividade em rede (os meios de transporte e roteamento das informações). Segundo Kwet, “colonização” significa precisamente a posse estruturante dessa infraestrutura digital, que se estende a uma série de subsistemas dependentes – comunicação, entretenimento, trabalho, transporte, educação e outros. O colonialismo digital manifesta-se, assim, na restrição das possibilidades de renda dos países e populações “colonizados”, na sua extrema dependência dos meios e recursos controlados pelo Norte Global, na imposição de uma divisão internacional do trabalho que favorece os países centrais, na manutenção de hegemônias de classe, na captura de mecanismos de ensino, formação e profissionalização, e, por fim, na administração e controle de múltiplas formas de interação social através da captura de políticas públicas, do planejamento econômico, dos serviços de inteligência e das práticas de policiamento.

Essas discussões sobre o colonialismo digital têm acontecido em um contexto que Couldry e Mejias (2023) denominam um *decolonial turn in data and technology research*, em texto em que se propõe a fazer uma genealogia do conceito de colonialismo digital. Eles lembram que, paralelamente às primeiras formulações sobre *data colonialism*, outros autores já mobilizavam lentes coloniais para analisar o digital no Sul Global. Entre esses trabalhos, destacam-se aqueles que tematizam as infraestruturas e os regimes de dependência tecnológica, como Thatcher *et al.* (2016), que aparecem como um marco inicial ao introduzir a noção de colonialismo de dados. Em

2019, Stefania Milan e Emiliano Treré, acadêmicos do sul da Europa, organizaram um dossiê intitulado *Big Data from the South*, na revista *Television & New Media*. Ainda que nem todos os textos adotassem diretamente a tese colonial, alguns trabalhos – incluindo o de Couldry e Mejias e, especialmente, o de Paula Ricaurte (2019) – ampliaram esse campo conceitual. Ricaurte desenvolve uma leitura explicitamente decolonial, interseccional e feminista do conceito de colonialismo de dados como articulado por Couldry e Mejias, apontando as continuidades entre a violência epistêmica histórica e as práticas contemporâneas de coleta, análise e exploração de dados.

Cabe sublinhar também, na literatura internacional sobre estudos críticos de dados (*critical data studies*), as menções à ideia de justiça de dados (*data justice*). Linnet Taylor (2017) formula o conceito como a busca por “justiça na forma como as pessoas são visibilizadas, representadas e tratadas como resultado de sua produção de dados digitais”. A potencial subalternidade de grupos sociais fica evidente quando ela discute um dos pilares da justiça de dados, o qual é o engajamento com as tecnologias. Segundo ela, essa questão se cruza com a teoria pós-colonial, enfatizando como os dados são processados e analisados nos mercados nacionais e globais, posiciona os indivíduos como subalternos em relação àqueles que controlam o processamento dessas informações. Ela interliga esse debate com as discussões em torno dos obstáculos que os sistemas de dados podem significar para o florescimento humano, sendo importante na elaboração desse enquadramento um olhar para um ecossistema global que vai além das fronteiras dos países. Lyon (2021), por sua vez, enquadra o florescimento humano como algo que abrange mais do que agência ou autonomia individual, enfatizando suas dimensões inerentemente relacionais e coletivas. O florescimento humano depende de condições sociais e institucionais que não podem se sustentar em ambientes marcados por vigilância assimétrica e reprodução ou intensificação de hierarquias de poder e classificatórias. Lyon aponta que a transparência deve ser, em especial, direcionada para aqueles que controlam os sistemas de vigilância. Justiça

de dados e florescimento humano aparecem como componentes inseparáveis de uma crítica mais profunda às estruturas globais que governam a vida digital.

Assim, podemos sublinhar como a “virada decolonial digital” (*digital decolonial turn*) proposta por Casili é condição necessária para uma ideia de florescimento humano no digital, um conceito eminentemente marcado pela interdependência entre grupos sociais diversos. O foco de Casili está nas relações de trabalho no digital, que ele procura visibilizar como uma categoria ampla, indo muito além do trabalho pago ou simplesmente mal pago, abrangendo, por exemplo, atividades dos “usuários” das plataformas, os quais contribuem na produção de um valor monetário por elas apropriado. Casili marca explicitamente como é necessário reconhecer que o trabalho nas plataformas digitais é consideravelmente feito por aqueles na “base da pirâmide”, estando igualmente envolvidos nos mecanismos de extração de dados por meio do labor digital. Nessa análise, introduz ainda o conceito de colonialidade digital, construído a partir de autores latino-americanos como Aníbal Quijano (2000) e Nelson Maldonado-Torres (2007).

Quijano (2000) propõe o conceito de colonialidade do poder, articulando as relações coloniais de dominação à expansão da acumulação capitalista na globalização. A colonialidade do poder manifesta-se na classificação da população mundial com base em critérios raciais, culturais e étnicos, criando estruturas institucionais destinadas a governar esses grupos e assegurar sua subalternização por meio da inferiorização permanente de seus modos de ser, existir e produzir conhecimento. É difícil não perceber como essas classificações históricas e estruturais acabam sendo embutidas – e muitas vezes ampliadas – pelos sistemas digitais contemporâneos, que operam classificações sociais automatizadas baseadas em vigilância. Maldonado-Torres (2007), por sua vez, desloca o foco para as experiências subjetivas ao formular a colonialidade do ser, indicando que os sujeitos modernos, herdeiros do colonialismo histórico em posições de poder ou subalternos, compartilham traços existenciais específicos. Nesses sujeitos, a expe-

riência da existência já está atravessada pela lógica colonial, marcada por hierarquizações que atribuem humanidade plena a uns e negam-na a outros.

Deste modo, por meio das diversas facetas da experiência de colonialidade, ideais como o de florescimento humano acabam sendo entendidos como um privilégio das populações do Norte Global, na contramão do que propõe Lyon e Taylor. Cabe aqui explicitar o nosso entendimento das categorias de Norte e Sul Global, conceitos que muitas vezes são amarrados à sua dimensão geopolítica – que é relevante, mas que não dá conta da complexidade histórica das relações sociais contemporâneas. Como observa Mahler (2017), Sul Global deixou de designar apenas Estados-nação economicamente desfavorecidos, passando a nomear pessoas e espaços negativamente impactados pela globalização capitalista e, podemos acrescentar, pela colonialidade. Essas populações frequentemente estão do sul geográfico. Mas, do mesmo modo, é necessário considerar que também existem populações vulnerabilizadas em países ricos, vivendo condições precárias que as aproximam das experiências do Sul Global. São imigrantes, refugiadas ou pertencem a grupos racializados e classes precarizadas. Importante ressaltar que são justamente essas populações que, na economia de plataformas, geralmente ocupam os piores postos de trabalho, como entregadores, trabalhadores em fazendas de cliques, motoristas de aplicativo etc. Simetricamente, há grupos privilegiados em países do Sul que, por seus corpos, histórias e posições sociais, se alinham mais ao Norte Global em termos de recursos, práticas e imaginários. Voltaremos a isso, mas são justamente esses grupos os intermediários que atuam na facilitação da presença das grandes plataformas no Sul Global.

Portanto, Norte e Sul Global não descrevem apenas lugares, mas posições estruturais dentro de um sistema de desigualdades históricas produzido e atualizado pela colonialidade. Empregar esses termos é uma opção sociológica – ao destacar processos de dominação socioeconômica e materialidades ecológicas – e também política, ao assumir uma perspectiva posicionada para a escolha de categorias analíticas e

de processos sociais a serem questionados. O que nos leva a indagar, de que ética se fala e que ética é possível se estamos lidando com estruturas materiais e simbólicas de extração de dados e recursos naturais, exploração de trabalho, epistemi-cídio e de atualização e intensificação de estruturas históricas de classificação entre aqueles com direito ao florescimento e outros não reconhecidos como humanos?

SOBERANIA DIGITAL COMO RESPOSTA À COLONIALIDADE

É a partir das discussões sobre colonialismo e colonialidade digital que propomos entender a emergência das discussões sobre soberania digital no Brasil. O país possui comunidade acadêmica e sociedade civil bastante engajadas, pelo menos desde o início dos anos 2000, em temas como inclusão digital; *software* livre e aberto; e direitos digitais, com ênfases distintas a depender do período tecnopolítico em análise. Esses grupos têm convergido para o tema soberania digital em parte como resposta à percepção de que as grandes plataformas digitais subordinam seus usuários a sistemas que oferecem recursos fechados e não-criativos; apropriam-se e privatizam conhecimentos comuns que deveriam ser compartilhados; e violam direitos, em especial de minorias, por meio de práticas de moderação injustas e não transparentes.

No continente europeu, as menções a soberania digital surgem em documentos oficiais da União Europeia desde meados dos 2000, mas é uma preocupação que ganha tração progressivamente a partir de 2013, com as revelações de Edward Snowden sobre o esquema global de espionagem capitaneado pelos EUA, e se intensifica no fim da década, a partir da constatação de perda de competitividade da empresas europeias frente as dos EUA (Pohle; Thiel, 2020).

Ainda que o movimento *software* livre latino americano na década de 2000 articulasse fortemente ideais de autonomia técnica, independência tecnológica e resistência às formas hegemônicas de controle e dependência (Evangelista, 2010), o vocabulário político da época não formulava a noção de “soberania digital”, que só emergiria mais tarde, ainda que haja uma conexão clara entre os dois momentos.

O primeiro documento brasileiro a mencionar “soberania digital” em termos de reivindicação política é a carta, endereçada ao então candidato a presidência, Luis Inácio Lula da Silva, em meados de 2022, intitulada “Programa de Emergência para a soberania digital” (CARTA SOBERANIA DIGITAL, 2022). Escrito e assinado por centenas de acadêmicos e ativistas, o documento alude a pontos convergentes ao tripé de Michel Kwet que conformariam o colonialismo digital – cabos, *hardware* e *software* –, além de mencionar a extração de dados para operações de vigilância. São mencionados na carta o sucateamento e franqueamento de estruturas nacionais a empresas estrangeiras (*big techs*), a extração de dados e a submissão da economia do país a um papel de consumidora de tecnologias. Pedem uma agenda de investimentos em tecnologia digital, afirmando que

a agenda de transformação digital brasileira precisa ser orientada ao bem-estar e ao florescimento humano de seus cidadãos, assim como ao enfrentamento da desigualdade, do racismo algorítmico e dos novos meios de segregação.

As ambições vão além da atuação soberana do Estado, no sentido geopolítico, afirma-se que é preciso “colocar no centro de nossa proposta a expansão da tecnodiversidade, da luta contra o epistemicídio, do desbloqueio da inventividade de nosso povo”.

No fim da década de 2010, acadêmicos ligados aos estudos das tecnologias digitais na educação e da vigilância passaram a denunciar os acordos que as universidades brasileiras progressivamente foram estabelecendo com grandes empresas de tecnologia dos EUA. Sem ouvir a comunidade acadêmica de forma consistente, aprovando documentos apressadamente, consultando apenas seu corpo técnico, reitores das principais universidades assinaram acordos de uso “gratuito” de serviços das plataformas digitais, como e-mail e armazenamento, desmobilizando as estruturas próprias das instituições. A carta cita explicitamente os estudos do Observatório Educação Viggiada, que informava que

79% das Instituições Públicas de Ensino Superior do país têm seus e-mails institucionais alocados em servidores privados, localizados fora do país, e que são gerenciados por empresas envolvidas no lucrativo mercado de coleta, análise e comercialização de dados pessoais.

Evangelista e Gonsales (2024), a partir de novos dados publicados em estudo do CGI.br (2023) e de uma análise detalhada de parcerias firmadas por universidades e secretarias estaduais, mostram que esse processo não ocorreu de forma espontânea, mas foi facilitado por intermediários locais – gestores públicos, consultores, técnicos e lideranças educacionais que atuaram como pontes para a entrada das grandes plataformas no sistema educacional brasileiro. Trata-se do fenômeno mencionado anteriormente: segmentos sociais que vivem no Sul Global, mas cujos interesses, imaginários e práticas estão mais alinhados ao Norte Global do que ao contexto em que estão inseridos.

O principal articulador do documento, professor da UFABC, Sergio Amadeu, deu repetidas declarações de que a ideia de soberania digital, contida na carta e em documentos posteriores, evocava o termo em analogia à reivindicação por “soberania alimentar”, pelo Movimento dos Trabalhadores Sem Terra (MST) e pela Via Campesina.

O movimento campesino luta pela terra, pela reforma agrária, mas também quer ter soberania alimentar. O que significa isso? Ele quer decidir o que plantar e o que comer, sem a dependência desses grandes grupos que controlam os alimentos (SBPCNET, 2025).

Vale lembrar que, em 2005, o Fórum Internacional de *Software* Livre, organizou o “banco de sementes livres”, com o objetivo de oferecer às comunidades indígenas e quilombolas do Rio Grande do Sul sementes livres de modificações genéticas e sobre as quais não incida nenhum direito de propriedade intelectual. O conceito de soberania alimentar foi expresso pela primeira vez em documento da Via Campesina, em 1996 (Haché, 2014). Mas foi sendo aperfeiçoado ao longo dos anos, até ser descrito em documentos da sociedade

civil, que chegaram a ser usados em cursos do então Ministério do Desenvolvimento Social e Combate à Fome (MDS), da seguinte forma:

importam à soberania alimentar a autonomia e as condições de vida e de trabalho dos agricultores familiares e camponeses, o que se reflete na produção de alimentos de qualidade, seguros, diversos, ambientalmente sustentáveis e adequados à cultura local. Esse conceito é também relevante no que diz respeito à soberania das nações e sua autossuficiência com relação aos alimentos para consumo interno. Remete, ainda, à preservação de sementes tradicionais (crioulas) e da biodiversidade agrícola, além da valorização de cultura e hábitos alimentares de diversas populações (Leão, 2013).

Ainda assim, talvez por trazer um termo (soberania) tradicionalmente ligado a discussões da ciência política referentes à autoridade política de Estados, diversos outros grupos sentiram a necessidade de complementar ou retrabalhar o termo. O capítulo brasileiro da Internet Society, em relatório de pesquisa feito com a Fundação Getúlio Vargas, afirma que existiria uma “variedade de interpretações e narrativas” relacionadas ao termo soberania digital e manifesta preocupação com relação aos efeitos sobre a governança multissetorial, apontando que “encontrar um equilíbrio entre a soberania nacional e a cooperação internacional é complexo” (Camelo, 2024). Já o Coletivo de Tecnologia e Soberania Digital do PSOL procurou adicionar a palavra popular ao termo soberania digital. Formulou um decálogo da soberania digital popular, mas descreveu o termo da seguinte maneira (a ênfase é dos autores originais):

Para nós, a Soberania Digital Popular, ou soberanias digitais populares, se trata da construção de ferramentas, infraestruturas e plataformas socialistas e livres, a criação de condições para tal, bem como a elaboração de um imaginário coletivo a partir de análises conjunturais que **viabilize o rompimento do colonialismo digital e de dados**. Isto é, a soberania digital popular

se baseia na defesa do desenvolvimento das condições materiais e subjetivas para apropriação popular da sociedade da informação, de forma a promover autonomias científica e tecnológica orientadas ao bem comum e não que promovam a dominação social, política, econômica, cultural e geopolítica. Em síntese, a soberania digital popular se refere à criação e à adoção de tecnologias – digitais ou não – emancipadoras a partir da sabedoria popular e da ciência cidadã que sejam facilitadoras da paz, da autodeterminação dos povos, da promoção das igualdades digitais, da proteção dos direitos humanos e do meio ambiente, portanto, socialistas e livres (PSOL, 2023).

CONCLUSÃO

O levantamento do contexto acadêmico, histórico e político que levou à formulação e à adoção do termo soberania digital como reivindicação política permite reconhecer que ele constitui uma categoria analítica mais rigorosa e produtiva do que noções vagas de “ética” largamente mobilizadas no debate público sobre tecnologia. Se a ética aparece, com frequência, como um significante esvaziado – utilizado sobretudo por setores empresariais como dispositivo discursivo para evitar regulações, diluir responsabilidades ou deslocar para o plano individual problemas estruturais – a soberania digital emerge, ao contrário, de diagnósticos situados de processos que marcam a tecnopolítica no Sul Global.

A partir de um posicionamento situado, no sentido proposto por Haraway (1995), afirmamos que o termo soberania digital, por mais que pareça polissêmico quando observado de uma distante falsa neutralidade, adquire contornos precisos quando analisado a partir de sua história de emergência no Brasil e na América Latina. Ele não nasce do setor empresarial – nacional ou internacional – tampouco de iniciativas tecnocráticas de governos, embora a estes sejam dirigidas demandas centrais. O termo é produzido, e assim deve ser entendido seu sentido, sobretudo por ativistas, pesquisadores, organizações da sociedade civil, movimentos sociais e comunidades técnicas que vivem cotidianamente os efeitos

da colonialidade digital: dependência tecnológica, vigilância assimétrica, extrativismo de dados, imposição de plataformas estrangeiras, erosão das infraestruturas públicas e apagamento sistemático das capacidades criativas e inventivas dos povos do Sul.

Reconhecer essa história é fundamental porque permite compreender a soberania digital não como uma aspiração de maior controle estatal sobre redes e dados, mas como uma categoria política articulada a noções de justiça de dados, florescimento humano e tecnodiversidade. Trata-se de reivindicar não apenas a proteção de infraestruturas, mas a possibilidade de construir um futuro digital que não reencene os modos históricos de dominação que marcaram – e continuam marcando – as relações entre Norte e Sul. É por isso que a analogia com a soberania alimentar não é mero recurso retórico, mas uma operação conceitual: ambas reivindicam condições materiais e institucionais para que populações historicamente vulnerabilizadas tenham controle sobre os meios que sustentam sua vida, sua cultura e sua autonomia coletiva.

Nesse sentido, a soberania digital, pensada desde o Sul Global, recoloca no centro do debate sobre tecnologia aquilo que a retórica empresarial da “ética” obscurece: que não haverá florescimento humano possível enquanto persistirem estruturas de colonialidade que organizam o trabalho, a vigilância, os dados, as classificações e o acesso às infraestruturas. Por isso, soberania digital é uma resposta situada que vai além dos limites da ética empresarial; é uma categoria que reconhece que as injustiças do ecossistema digital não são desvios ou danos colaterais, mas atualizações de uma longa história de hierarquização e apropriação.

REFERÊNCIAS

- BAKER-BRUNNBAUER, J. Management perspective of ethics in artificial intelligence. *AI & Ethics*, v. 1, n. 2, p. 173-181, 2021. DOI: 10.1007/s43681-020-00022-3.
- CAMELO, Ana Paula *et al.* *Soberania digital* [recurso eletrônico]: para quê e para quem? Análise conceitual e política do conceito a partir do contexto brasileiro. São Paulo: FGV Direito SP, 2024.

CARTA SOBERANIA DIGITAL. *Carta pela Soberania Digital*. 2022. Disponível em: <https://cartasoberaniadigital.lablivre.wiki.br/carta/>. Acesso em: 26 jan. 2026.

CASILLI, Antonio A. Digital labor studies go global: toward a digital decolonial turn. *International Journal of Communication*, Los Angeles, v. 11, p. 3934–3954, 2017.

COMITÊ GESTOR DA INTERNET NO BRASIL (CGI.br). *Educação em um cenário de plataformação e de economia dos dados: parcerias e assimetrias*. Coordenação editorial: Rafael de Almeida Evangelista; pesquisa e redação: Priscila Gonsales. São Paulo: CGI.br, 2023. Disponível em: <https://cgi.br/publicacao/educacao-em-um-cenario-de-plataformizacao-e-de-economia-dos-dados-parcerias-e-assimetrias>. Acesso em: 26 jan. 2026.

COULDRY, Nick; MEJIAS, Ulises A. *The costs of connection: how data is colonizing human life and appropriating it for capitalism*. Stanford: Stanford University Press, 2019.

CRAWFORD, Kate. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven: Yale University Press, 2021.

DIJCK, Jose van. Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society*, v. 12, n. 2, p. 197–208, 9 maio 2014.

EVANGELISTA, Rafael. *Traidores do movimento: política, cultura, ideologia e trabalho no software livre*. 2010. Tese (Doutorado em Antropologia Social) – Universidade Estadual de Campinas, Campinas, 2010.

EVANGELISTA, Rafael; GONSALES, Priscila. A plataformação da educação no Sul Global e seus laços com os atores do capitalismo de vigilância. In: ALVES, Lynn Rosalina Gama; LOPES, David Santana (org.). *Educação e plataformas digitais: popularizando saberes, potencialidades e controvérsias*. Salvador: EDUFBA, 2024. p. 17–38. Disponível em: <https://repositorio.ufba.br/bitstream/ri/39372/3/Educa%C3%A7%C3%A3o%20e%20plataformas%20digitais-digital.pdf>. Acesso em: 26 jan. 2026.

HACHÉ, Alex. Soberania tecnológica. In: HACHÉ, Alex (Org.). *Soberania tecnológica*. [S.l.: s.n.], 2014. Disponível em: https://archive.org/details/dossier-st-cast-2014-06-30_202105/Libro_soberania_tecnologica_cast_final Acesso em: 21 nov. 2025.

- HARAWAY, Donna J. Saberes localizados: a questão da ciência para o feminismo e o privilégio da perspectiva parcial. *Cadernos Pagu*, Campinas, v. 5, p. 7–41, 1995.
- KWET, Michael. *Digital degrowth: technology in the age of survival*. London: Pluto Press, 2024.
- LEÃO, Marília (Org.). *O direito humano à alimentação adequada e o sistema nacional de segurança alimentar e nutricional*. Brasília: ABRANDH, 2013.
- LYON, David. Surveillance, transparency, and trust. In: *Trust and Transparency in an Age of Surveillance*. [S.l.]: Routledge, 2021. p. 243–257.
- MAHLER, Anne Garland. Global South. *Oxford Bibliographies*. Oxford: Oxford University Press, 2017. Disponível em: <https://www.oxfordbibliographies.com/>. Acesso em: 26 jan. 2026.
- MALDONADO-TORRES, Nelson. On the coloniality of being: contributions to the development of a concept. *Cultural Studies*, Londres, v. 21, n. 2–3, p. 240–270, 2007.
- NYABOLA, Nanjala. Sovereign identity crisis: What does a decolonial approach to digital technology reveal about the sovereign identity crisis? *Temple Law Review*, v. 95, n. 4, p. 593–603, 2023.
- POHLE, Julia; THIEL, Thorsten. Digital sovereignty. *Internet Policy Review*, v. 9, n. 4, 16 dez. 2020. Disponível em: <https://policyreview.info/concepts/digital-sovereignty>. Acesso em: 21 nov. 2025.
- PSOL – *Tecnologia e Soberania Digital*. Decálogo para a Soberania Digital Popular. [S.l.: s.n.], 2023. 1 vol. Disponível em: https://psoltecnologia.com.br/wp-content/uploads/2023/09/PSOL_Decalogo_SoberaniaDigitalPopul ar_2023.pdf. Acesso em: 21 nov. 2025.
- QUIJANO, Anibal. Coloniality of power, Eurocentrism, and Latin America. *Nepantla: Views from the South*, Durham, v. 1, n. 3, p. 533–580, 2000.
- RICAURTE, Paola. Data epistemologies, the coloniality of power, and resistance. *Television & New Media*, Londres, v. 20, n. 4, p. 350–365, maio 2019.
- SBPCNET – SOCIEDADE BRASILEIRA PARA O PROGRESSO DA CIÊNCIA. Brasil só terá soberania nacional quando tiver a independência de seus dados. 2025. Disponível em: <https://portal.sbpcnet.org.br/noticias/brasil-so-tera-soberania-nacional-quando-tiver-a-independencia-de-seus-dados/>. Acesso em: 26 jan. 2026.

TAYLOR, Linnet. What is data justice? The case for connecting digital rights and freedoms globally. *Big Data & Society*, Londres, v. 4, n. 2, p. 1–14, dez. 2017.

THATCHER, Jim; O’SULLIVAN, David; MAHMOUDI, Dillon. Data colonialism through accumulation by dispossession: new metaphors for daily data. *Environment and Planning D: Society and Space*, Londres, v. 34, n. 6, p. 990–1006, 2016.

ZUBOFF, Shoshana. *The age of surveillance capitalism: the fight for a human future at the new frontier of power*. Nova Iorque: PublicAffairs, 2019.

**SIMBIOSE HUMANO-MÁQUINA COMPUTACIONAL:
O LEGADO TECNOLÓGICO VISIONÁRIO DE LICKLIDER**

**HUMAN-COMPUTATIONAL MACHINE
SYMBIOSIS: THE VISIONARY TECHNOLOGICAL
LEGACY OF LICKLIDER**

DOI: 10.29327/5798915.1-7

*Walter Teixeira Lima Junior*¹⁶

RESUMO: O estadunidense Joseph Carl Robnett Licklider é reconhecido por moldar as bases da computação pessoal e da internet. Sua contribuição seminal, a Simbiose Humano-Máquina Computacional (proposta em 1960), era uma visão científica surpreendente para a época dos grandes *mainframes* e processamento em lote. Licklider propôs uma relação simétrica e cooperativa, tomando o conceito de simbiose da biologia para configurar sistemas que ampliam mutuamente as capacidades. Essa abordagem se diferencia da visão tradicional, que enxergava a máquina como uma extensão mecânica do homem, denominada de relação *master-slave*, influenciada pela Revolução Industrial e Cibernética. A simbiose se fundamenta na complementaridade: o poder intelectual efetivo da parceria superaria amplamente o de cada componente isolado. A divisão de trabalho estabelecia que o humano definiria os objetivos e hipóteses, enquanto a máquina executaria o trabalho rotineiro/algorítmico, como calcular e armazenar grandes volumes de informação. O legado de Licklider, que buscava libertar os humanos das tarefas enfadonhas para atividades criativas, permanece o paradigma filosófico-tecnológico que sustenta a pesquisa moderna, incluindo a Inteligência Artificial Generati-

¹⁶ Docente do Programa de Pós-graduação Profissional em Inovação Tecnológica da Universidade Federal de São Paulo (Unifesp) e líder do Grupo de Pesquisa Sistemas Cognitivos Artificiais e Robótica Social, walter.lima@unifesp.br.

va (*Foundation Models*). Eticamente, a visão de Licklider impõe que a tecnologia deve ser construída para aumentar a capacidade humana, e não para substituí-la ou aprisioná-la.

Palavras-chave: Simbiose Humano-Máquina Computacional; Inteligência Artificial Generativa; Filosofia da Tecnologia; Interatividade; Interdisciplinaridade.

Abstract: The American Joseph Carl Robnett Licklider is recognized for shaping the foundations of personal computing and the internet. His seminal contribution, “Man-Computer Symbiosis” (proposed in 1960), was a remarkable scientific vision for the era of large mainframes and batch processing. Licklider proposed a symmetrical and cooperative relationship, borrowing the concept of symbiosis from biology to configure systems that mutually amplify capabilities. This approach differs from the traditional view, which perceived the machine as a mechanical extension of man—termed a master-slave relationship—influenced by the Industrial Revolution and Cybernetics. Symbiosis is grounded in complementarity: the effective intellectual power of the partnership would vastly surpass that of each isolated component. The division of labor established that the human would define the goals and hypotheses, while the machine would execute the routine/algorithmic work, such as calculating and storing large volumes of information. Licklider’s legacy, which sought to free humans from tedious tasks for creative activities, remains the philosophical-technological paradigm underpinning modern research, including Generative Artificial Intelligence (*Foundation Models*). Ethically, Licklider's vision dictates that technology must be built to augment human capacity, rather than to replace or imprison it.

Keywords: Human-Computer Symbiosis; Generative Artificial Intelligence; Philosophy of Technology; Interactivity; Interdisciplinarity.

INTRODUÇÃO

O estadunidense Joseph Carl Robnett Licklider é considerado como “Johnny Appleseed¹⁷ da Computação”, devido a sua capacidade de semear ideias e executar projetos audaciosos. Licklider foi um visionário que moldou as bases da computação pessoal e da internet. Suas ideias inovadoras sobre a interação entre humanos e máquinas criaram um paradigma que transformou a forma como interagimos com a tecnologia.

Reconhecido como um pioneiro no desenvolvimento de reflexões sobre a relação entre humanos e computadores, Licklider foi além da simples oferta de dados por sistemas computacionais (Fano, 1998). Essa visão científica foi surpreendente na década de 1960, especialmente considerando o estado da tecnologia computacional da época – isto é, grandes *mainframes*, cartões perfurados e processamento em lote (*batch processing*) (Greitzer; Griffith, 2006).

A esperança é que, em poucos anos, cérebros humanos e máquinas computacionais estejam fortemente interligados, e que essa parceria resultante pense de uma forma como nenhum cérebro humano jamais pensou e processe dados de um modo nunca antes alcançado pelas máquinas de processamento de informação que conhecemos hoje (Licklider, 1960).

Licklider avança nas discussões e experimentos sobre Interação Humano-Computador (*Human Computer Interaction*), propondo uma nova forma de relação entre homem e computador: a simbiose. Ele tomou o conceito de simbiose da biologia. Essa abstração implica uma relação simétrica entre o humano e o computador, configurando sistemas que ampliam as capacidades e beneficiam mutuamente dois ou mais organismos, biológicos ou sintéticos.

Seu objetivo era aprimorar a relação simbiótica entre os parceiros humano e computador, necessária para auxiliar a atividade intelectual (Baron, 2011). Segundo Licklider, para

¹⁷ Johnny Appleseed é uma figura lendária estadunidense, baseada na vida real de John Chapman, conhecido por plantar macieiras pelo meio-oeste dos EUA, sendo um semeador ideais de paz e conservação.

melhorar o desempenho na relação entre homem e computador, o computador deve ser um participante ativo, e não apenas um dispositivo passivo de comunicação (Adel; Brangier, 2013). Ele previu que, em poucos anos, os seres humanos seriam capazes de se comunicar de forma mais eficaz por meio de uma máquina do que face a face, conforme escreveu em um artigo com Taylor (Licklider; Taylor, 1968), em uma época em que os computadores eram usados quase exclusivamente por cientistas.

Como precursor, ele antecipava a futura evolução da computação, que deixaria de ser apenas uma máquina para tornar-se uma parceira capaz de superar limitações humanas em uma multiplicidade de áreas. Para descrever essa relação intensa, Licklider utilizou a metáfora da simbiose para qualificar o vínculo entre humanos e tecnologia (Adel; Brangier, 2013).

O conceito de simbiose de Licklider também foi mencionado por outro cientista importante no estudo da relação humano-máquina para cooperação mutuamente benéfica: Douglas Engelbart. Engelbart analisou a viabilidade de construir uma superestrutura como uma extensão sintética da estrutura natural humana. Ao ampliar nossa inteligência, segundo ele, estamos aplicando o princípio da estruturação sinérgica que foi seguida pela evolução natural no desenvolvimento das capacidades humanas básicas. O que fizemos no desenvolvimento de nossos meios de ampliação foi construir uma superestrutura que é uma extensão sintética da estrutura natural na qual ela se baseia. Em um sentido muito real, representado pela evolução contínua de nossos meios de ampliação, o desenvolvimento da inteligência artificial vem ocorrendo há séculos (Engelbart, 1962).

A ideia central é como fazer um computador e uma pessoa pensarem juntos, compartilhando e dividindo a carga – algo que fundamenta a simbiose de Licklider, principalmente no contraste entre o heurístico e o algorítmico (Aspray, 1988). Quase sessenta anos após Licklider e Engelbart formularem suas visões, a maior parte dos impedimentos para uma relação frutífera com a máquina – isto é, barreiras técnicas ou econômicas – desapareceu (Schalk, 2008, p. 2).

Hoje, novos termos surgiram tentando demonstrar que as tecnologias atuais podem ser usadas em uma relação mútua entre homem e computador. Na era das *Foundation Models*, via a denominada Inteligência Artificial Generativa, nós criamos a tarefa, e o computador a executa e nos apresenta os resultados.

NA CONTRAMÃO DA TECNOLOGIA SUBORDINADA ÀS INTENÇÕES HUMANAS

O campo de Interação Humano-Computador (HCI) vem sendo pesquisado cientificamente ao longo dos últimos sessenta anos, embora sua primeira menção acadêmica tenha aparecido apenas em 1975. A HCI deriva de teorias sobre o comportamento humano ao utilizar máquinas computacionais. Muitas dessas teorias baseiam-se na visão da tecnologia como uma extensão mecânica do homem. Tipicamente, computadores são entendidos como ferramentas destinadas a tornar as pessoas mais produtivas em seus trabalhos (Griffith, 2005).

A tecnologia computacional serve aos propósitos humanos; em outras palavras, a tecnologia se subordina às intenções humanas. Essa visão tem como objetivo fazer com que as máquinas auxiliem o ser humano na tomada de decisão e na execução de tarefas. A ideia da máquina como extensão do homem tem origem na Revolução Industrial, período em que ocorreu a transição para novos processos de manufatura baseados em máquinas mecânicas. Na Revolução Industrial, as máquinas representavam a extensão do trabalho humano (relação mestre-servo). A cibernética também foi influenciada por essa forma de pensar ao criar sistemas de controle que consideravam a assimetria entre humano e máquina, nos quais o foco dos fatores humanos era a usabilidade (Griffith, 2005).

Essa é uma visão tradicional da interação humano-máquina, que considera apenas o fator de comunicação facilitada e ergonomicamente otimizada. Máquinas e sistemas computacionais são produzidos para cooperar na busca de soluções para problemas formulados. Assim, a relação entre máquina e humano é uma relação na qual a máquina se subordina aos desejos humanos.

Contrária a essa visão tecnológica, as reflexões de Joseph Carl Robnett Licklider direcionavam-se à cooperação mútua entre homem e computador – denominada simbiose homem-computador. A visão de Licklider permanece lógica e relevante atualmente. Seu trabalho é reconhecido como visionário ao propor um relacionamento simbiótico entre humanos e computadores. Como conceito, a simbiose homem-computador difere de maneira importante do que North chamou de “homem mecanicamente estendido” (Licklider, 1960).

Essa visão disruptiva tecnologicamente é fruto do conhecimento interdisciplinar adquirido e cultivado por Licklider, o qual construído ao longo de anos de estudo em ciências aplicadas, ciências naturais e ciências sociais. As conexões entre essas áreas permitiram a Licklider propor uma cooperação entre computador e ser humano, diferente da visão tradicional de interação humano-máquina, que considera apenas a otimização ergonômica da comunicação. Licklider percebeu que o computador e o ser humano podem trabalhar juntos. *Man-Computer Symbiosis* aborda principalmente ideias sobre como fazer o computador e a pessoa pensarem em conjunto, compartilhando e dividindo a carga – sobretudo no contraste heurístico *versus* algorítmico.

Entretanto, o pensamento interdisciplinar de Licklider começou a se estruturar academicamente em 1937, quando obteve três diplomas de bacharel pela Washington University, nas áreas de física, matemática e psicologia – uma combinação rara que o tornou um cientista interdisciplinar. Em seus estudos sobre psicoacústica, Licklider aproximou-se da biologia. Em 1951, publicou o artigo “Duplex Theory of Pitch Perception”, resultado de sua pesquisa de doutorado na University of Rochester, em 1942, considerado a base para modelos modernos na área de percepção sonora.

Seu pensamento interdisciplinar iniciado nos anos 1940 encontrou nos encontros multidisciplinares promovidos por Nobert Wiener, nos anos 1950, o ambiente propício para o desenvolvimento de suas ideias avançadas e projetos nos anos 1960.

SER HUMANO E MÁQUINA COMPUTACIONAL SE COMPLEMENTANDO

A visão de interação simbiótica entre ser humano e a máquina computacional digital, segundo Licklider, funcionaria no sentido da complementaridade:

Nós acreditamos que homem e computador se complementam, e que o poder intelectual de uma simbiose homem-computador efetiva irá ultrapassar amplamente o de cada componente isolado (Licklider; Clark, 1962).

Para essa complementariedade ser executada, Licklider elencou as características positivas de seres humanos e de máquinas computacionais neste processo de interação simbiótica. Assim, descreveu que o objetivo fundamental de projetar uma simbiose homem-computador é explorar essa complementaridade. A complementaridade surge das capacidades existentes entre os seres humanos e máquinas computacionais. Assim, Licklider formulou as áreas em que computador e humano desempenham tarefas melhor, a fim de criar um ambiente complementar:

- a. Definir metas e critérios – humano;
- b. Formular questões e hipóteses – humano;
- c. Selecionar abordagens – humano;
- d. Detectar relevância – humano;
- e. Reconhecer padrões e objetos – humano;
- f. Lidar com exigências imprevistas e de baixa probabilidade – humano;
- g. Armazenar grandes quantidades de informação – humano e computador;
- h. Recuperar informações rapidamente – humano e computador;
- i. Calcular rápida e precisamente – computador;
- j. Desenvolver progressivamente um repertório de procedimentos sem sofrer perdas por interferência ou falta de uso – computador (Licklider; Clark, 1962).

A complementaridade de Licklider pode ser baseada no conceito de sistemas acoplados, quando uma máquina pode ser acoplada a outra para formar uma única máquina.

O objetivo da “Simbiose Homem-Máquina” é, então, possibilitar o acoplamento confiável entre os humanos e suas informações “externas”, representadas por computadores digitais. Para alcançar esse acoplamento, foi necessário superar as barreiras de tempo e espaço, para que a simbiose pudesse operar como um processo único. Isso exigiu a invenção de ciclos de *feedback* com latência cada vez menor entre humanos e suas máquinas (Halpin, 2013).

A cibernética também entende que máquinas acopladas são de grande importância, porque cada máquina afeta a outra apenas ao afetar suas condições. Este é o mesmo princípio de Licklider para a simbiose homem-computador, onde dois organismos devem afetar um ao outro. Ashby (1957) afirmou que a complementaridade é uma propriedade fundamental das máquinas: elas podem ser acopladas. Duas ou mais máquinas completas podem ser acopladas para formar uma única máquina; e qualquer máquina pode ser vista como formada pelo acoplamento de suas partes, que podem ser entendidas como pequenas submáquinas. Esse acoplamento é de profunda importância científica, pois quando o experimentador conduz um experimento, ele está temporariamente acoplando-se ao sistema que estuda.

No entanto, Licklider criticou os argumentos de que algumas funções deveriam ser atribuídas a humanos e outras a máquinas. Ele classificou essa posição como “abordagem tradicional”, porque é particularmente insatisfatória: para que as funções principais envolvidas no trabalho com o corpo de conhecimento sejam realizadas de forma eficiente, é necessária uma ação sinérgica na qual humanos e máquinas participem juntos. A maioria dos esforços feitos durante a última década para definir “o que os humanos devem fazer” e “o que as máquinas devem fazer” falhou amplamente nesse ponto (Licklider, 1965).

Complementando essa aceitação, alguns estudos tentaram atribuir à tecnologia, ao usuário e ao contexto papéis

equivalentes na formação de uma relação especial – às vezes chamada acoplamento. Eles se concentraram no *feedback* e na co-construção entre os dois parceiros interativos. A noção de simbiose homem-computador (Licklider, 1960), inspirada na biologia, tem sido usada como ponto de partida, onde ambos simbioses – humano e tecnologia – beneficiam-se mutuamente de sua relação próxima. O objetivo dessa relação é combinar o melhor do humano e o melhor da tecnologia para ampliar as capacidades cognitivas humanas. Em sua visão, cada parceiro terá um papel na cooperação. Para Licklider, os usuários tenderão a preferir sistemas simbióticos que ampliem suas capacidades.

Embora máquinas operem por princípios muito diferentes do cérebro humano – porque os poderes e limitações de ambos são distintos –, juntas, a capacidade das máquinas e do ser humano se complementam, levando à possibilidade de que a combinação seja mais produtiva e poderosa do que cada um isoladamente (Norman, 1993).

No comentário de John Senders inserido no artigo *Subscribed Content Beyond Usability: The New Symbiosis* (Griffith, 2005), reforça-se a visão de que simbiose biológica envolve benefícios para ambos os membros da relação. Assim, usabilidade ótima não é simbiose. Uma conexão estreita entre usuários e computadores capaz de gerar benefícios para usuários é razoável – isso já existe quando existe usabilidade. Mas onde está o benefício para os computadores? Para que um computador se beneficie, ele deve ter metas e ser capaz de autoaperfeiçoamento ao tirar proveito da ligação com seres humanos (Griffith, 2005).

LICKLIDER E AS RAÍZES DAS FOUNDATION MODELS

A visão de Licklider tem a máquina computacional como uma extensão do cérebro, com o objetivo de buscar uma interação mais intuitiva entre a máquina e o cérebro. Nessa simbiose, o ser humano definiria os objetivos, formularia as hipóteses, determinaria os critérios e realizaria as avaliações (as tarefas heurísticas e formulativas). Já caberia à máquina computacional realizar o trabalho rotineiro (*routinizable*

work) necessário para preparar o caminho para *insights* e decisões (as tarefas algorítmicas) (Licklider; Taylor, 1968).

A meta de Licklider era libertar os humanos das tarefas rotineiras e enfadonhas, permitindo que eles se concentrassem em atividades criativas e inovadoras.

Mas, qual seria esse trabalho rotineiro? Dentro do pensamento de Licklider, mencionado acima, os itens g. Armazenar grandes quantidades de informação; h. Recuperar informações rapidamente; i. Calcular rápida e precisamente, seriam enquadrados nesse trabalho rotineiro.

Podemos ser apropriado reconhecer, neste ponto, que estamos a usar o termo “computador” para abranger uma ampla classe de máquinas de cálculo, processamento de dados e armazenamento e recuperação de informações. As capacidades das máquinas desta classe estão a aumentar quase diariamente. É, portanto, arriscado fazer declarações gerais sobre as capacidades da classe. Talvez seja igualmente arriscado fazer declarações gerais sobre as capacidades dos seres humanos. No entanto, certas diferenças genóticas na capacidade entre homens e computadores destacam-se, e elas têm uma influência sobre a natureza da possível simbiose homem-computador e o valor potencial de a alcançar (Licklider, 1960).

Assim, as máquinas computacionais devem facilitar o pensamento formulativo, ajudando os usuários a desenvolver e estruturar suas ideias de maneira mais eficaz. Isso implica não apenas em executar tarefas, mas em colaborar na criação de soluções inovadoras. A parceria simbiótica também visa permitir que humanos e computadores cooperem na tomada de decisões, especialmente em situações complexas onde a análise de dados e a criatividade humana são necessárias para chegar a conclusões eficientes.

Ao olharmos o desenvolvimento da tecnologia de *Foundation Model*, como modelo mestre ou uma base fundamental sobre a qual outras aplicações são construídas como as aplicações de Inteligência Artificial Generativa, percebe-se

que no seu cerne há semelhanças com as alusões contidas na Máquina de Turing e na visão de *Man-computer Symbiosis*, de Licklider, das máquinas computacionais “parecerem”, em certos fenômenos comportamentais e de linguagem, com o ser humano. Este “desejo” de tentar formalizar computacionalmente a linguagem humana possui décadas de desenvolvimento teórico e experimental.

O trabalho seminal de Shannon (1951), *Prediction and Entropy of Printed English*, estabeleceu as bases conceituais e teóricas que, décadas mais tarde, culminariam nos modernos *Foundation Models*, como o GPT-3 e outros modelos de linguagem de grande escala (*Large Language Models*).

As conexões entre o estudo de Shannon e a Inteligência Artificial Generativa atual são profundas, abrangendo desde a tarefa fundamental de previsão de texto até a maneira como esses sistemas lidam com a incerteza da linguagem. Em sua essência, o artigo de Shannon introduziu uma metodologia para estimar a entropia e a redundância da língua inglesa.

Para isso, ele propôs um experimento engenhoso: prever a próxima letra de um texto com base nas letras anteriores. Essa tarefa, aparentemente simples, revelou a riqueza de informações estatísticas contidas na linguagem e como o contexto influencia a previsibilidade.

Este ensaio científico de Shannon é citado em outro trabalho científico, *Long Range Constrains in The Statistical Structure of Printed English*, de Licklider e N.G. Burton, publicado em 1955. O ensaio focaliza na análise da redundância e previsibilidade da língua inglesa impressa, expandindo os conceitos da Teoria da Informação desenvolvidos por Shannon.

Outro encadeamento na busca da previsibilidade da próxima palavra, é o artigo *Trie Memory*, de 1960, de Edward Fredkin, Bolt Beranek e Newman. Este ensaio científico descreve um sistema de organização de memória inovador concebido para facilitar a recuperação de informações, visionando desenvolver programas de computador que podem ser conectados como as palavras e frases da fala para executar instantâ-

neamente a computação ou controle necessário. Este estudo utilizou como base o artigo do Shannon e do Licklider, citados acima. Que mostra uma rede de desenvolvimento e consolidação de um pensamento científico entre eles.

O artigo *Trie Memory* foi uma das bases para o pensamento seminal de J.C.R. Licklider sobre simbiose ser humano e máquina computacional, descrito no ensaio científico *Man-Computer Symbiose*.

A visão de parceria intelectual e de comunicação simétrica proposta por Licklider superou as limitações do modelo puramente técnico da Cibernética e se tornou o paradigma filosófico e tecnológico que sustenta a pesquisa moderna em *Cognitive Computing* e na interconexão global. Isso exige uma nova união entre Engenharia, Ciência da Computação, Ciências Cognitivas, Neurociências e Psicologia para concretizar totalmente a agenda de pesquisa proposta.

A conexão entre o pensamento seminal de J. C. R. Licklider sobre a Simbiose Homem-Máquina e as questões éticas que hoje cercam a Inteligência Artificial Generativa (IA Generativa) reside fundamentalmente na definição do papel do ser humano na parceria tecnológica e no acesso equitativo ao poder intelectual amplificado.

O foco ético de Licklider não era em proibições, mas sim em um imperativo de projeto: a tecnologia deve ser construída para aumentar a capacidade humana, e não para substituí-la ou aprisioná-la.

REFERÊNCIAS

- ADEL, Sonia; BRANGIER, Eric. Evolutions in the human technology relationship: rejection, acceptance and technosymbiosis. *IADIS International Journal on WWW/Internet*, 11(3):46-60, 2013.
- ASHBY, W. Ross. *An introduction to cybernetics*. second imp edition. London: Chapman & Hall Ltd, 1957.
- ASPRAY, William; KITCHER, Philip (ed.). *History and Philosophy of Modern Mathematics*. Minneapolis: University of Minnesota Press, 1988. 386 p. (Minnesota Studies in the Philosophy of Science, v. 11)

BARON, Shelly; ET AL. *A culture of innovation insider accounts of computing and live at BBN*. East Sandwich, MA: Waterside Publishing, 2011.

BURTON, N. G.; LICKLIDER, J. C. R. Long-Range Constraints in the Statistical Structure of Printed English. *The American Journal of Psychology*, Urbana, v. 68, n. 4, p. 650-653, Dec. 1955.

ENGELBART, Douglas C. Augmenting human intellect: conceptual framework. *Summary Report - Stanford Research Institute*, 49(3578):134, 1962.

FANO, Robert M. Joseph Carl Robnett Licklider, 1915–1990. In: NATIONAL ACADEMY OF SCIENCES. Biographical Memoirs. Volume 75. Washington, DC: The National Academies Press, 1998. p. 191-214.

FREDKIN, E. Trie memory. *Communications of the ACM*, p. 490-499, set. 1960.

GREITZER, Frank L.; GRIFFITH, Douglas. A Human-Information Interaction Perspective on Augmented Cognition. In *Augmented Cognition International Conference*, São Francisco, 2006.

GRIFFITH, Doug. Beyond Usability: The New Symbiosis. *Ergonomics in Design: The Quarterly of Human Factors Applications*, 13(3):30-34, June 2005.

HALPIN, Harry. Social Semantics. In *Social Semantics: the search for meaning on the web*, volume 13 of Semantic Web and Beyond, pages 9-50. Boston, MA: Springer US, 2013.

LICKLIDER, J. C. R. Man-Computer Symbiosis. *IRE Transactions on Human Factors in Electronics*, pages 4-11, 1960.

LICKLIDER, J. C. R. On-line man-computer communication. In *Proceedings of the May 1-3, 1962, spring joint computer conference on - AIEE-IRE '62 (Spring)*, pages 113-128, 1962.

LICKLIDER, J. C. R. *Libraries of the future*. Boston: MIT PRESS, 1965.

LICKLIDER, J. C. R.; TAYLOR, Robert W. The computer as a communication device. *Science and Technology*, (September):20-41, 1968.

NORMAN, Donald A. *Things That Make Us Smart: Defending Human At- tributes in the Age of the Machine*. MA: Addison-Wesley, 1993.

SCHALK, Gerwin. Brain-Computer Symbiosis. *J Neural Eng*, 5(1):1-28, 2008.

SHANNON, C. E. Prediction and entropy in printed English. *Bell Syst. Tech. J.*, v. 30, p. 50-64, 1951.

QUEM SOMOS?

BRUTUS ABEL FRATUCE PIMENTEL

Doutor em Filosofia pela Faculdade de Filosofia e Ciências Humanas da Universidade de São Paulo (FFLCH-USP). Professor do Departamento de Humanidades (IEFH) do Instituto Tecnológico de Aeronáutica (ITA), assessor do Programa de Atividade de Iniciação Científica e Tecnológica do ITA e membro do Grupo de Trabalho Equidade de Gênero do ITA. Atua nas seguintes áreas: Epistemologia, Filosofia da Ciência, Filosofia da Religião.

CV Lattes: <http://lattes.cnpq.br/1788437962199264>

CAROLINA VAITIEKUNAS PIZARRO

Professora Assistente Doutora no Departamento de Design e no Programa de Pós-Graduação em Design da Faculdade de Arquitetura, Artes, Comunicação e Design – FAAC – na Universidade Estadual Paulista, UNESP campus de Bauru. Lidera o Grupo de pesquisa “Design, Inteligência Artificial e impactos sociais” CNPq-UNESP criado em 2024 e coordena o LADIA – Laboratório Design, Inteligência Artificial e impactos sociais. É Doutora em Design pelo Programa de Pós-Graduação em Design na Faculdade de Arquitetura, Artes, Comunicação e Design da Universidade Estadual Paulista UNESP-Bauru (2019), Mestra em Design pelo mesmo programa (2014) e graduada em Desenho Industrial – Projeto de Produto na mesma Universidade (2010). Realiza investigações acerca da prática de projeto, da evolução do Design como área e do papel do designer como profissional frente às novas demandas da área e da sociedade. Seus interesses de pesquisa situam-se nas áreas de Design e Inteligência artificial; Inteligência artificial e seus impactos sociais; Prática de projeto e o papel do designer na sociedade; Teoria e crítica do design; Design centrado no ser humano.

CV Lattes: <http://lattes.cnpq.br/8867373608573768>

CASSIANO TERRA RODRIGUES

Professor de lógica, epistemologia e filosofia da ciência e coordenador do Laboratório de Filosofia, Lógica e Epistemologia da Tecno-Ciência (LabFILOETEC) no Instituto Tecnológico de Aeronáutica (ITA). Pesquisador colaborador do Centro de Estudos de Pragmatismo da PUC-SP, do Grupo de Pesquisa de Filosofia Medieval Latina e Filosofia Medieval em Árabe (FALSAFA) e da Rede Brasileira de Pesquisa em Semiótica Peirciana. Membro da Associação Filosófica Scientiae Studia, da Associação Brasileira de Estudos Semióticos e da Charles S. Peirce Society (Comitê Executivo (“member at large”). Sócio fundador da Sociedade Latino Americana Peirce. Presidente da Sociedade Brasileira de Lógica (2025-2027). Adepto da Pedagogia Corinthianista.

CV Lattes: <http://lattes.cnpq.br/3192898321927332>

DANIELA AMÉRICA DA SILVA

Instituto Tecnológico de Aeronáutica, ITA, DSc Eng. Eletrônica e Computação (ITA, 2022) e Mestre em TI (Univ. Melbourne, AUS, 2006). Especialista em Dados, pesquisadora (Ethics4AI/IDP & Mackenzie e IDEIA/ITA). Interesses de pesquisa: machine learning, processamento de linguagem natural, IA Ética. Publicações internacionais (AMCIS, IEEE RITA, MDPI Information). Revisora (SBSI, ISLA, SBC iSys).

CV Lattes: <http://lattes.cnpq.br/7901166170527255>

DELMO MATTOS DA SILVA

Professor de Filosofia do Departamento de Humanidades (IEFH), na Divisão de Ciências Fundamentais do Instituto Tecnológico de Aeronáutica (ITA), é também Professor Permanente do Programa de Pós-Graduação em Propriedade Intelectual e Transferência de Tecnologia para a Inovação (PROFNIT/ITA). Lidera o Grupo de Pesquisa Inovação e Dilemas Éticos da Inteligência Artificial (IDEIA/CNPq). Concluiu estágio de pós-doutorado em Teoria da Justiça pelo PPGDIR/UFMA (2017) e possui Doutorado (2008), Mestrado (2003, CAPES) e Bacharelado (2000) em Filosofia pela Universidade Federal do Rio de Janeiro. Coordena o GT Hobbes da ANPOF e integra a European Hobbes Society (desde

2020). Coordena o LabIDEIA (Laboratório de Pesquisa e Extensão Inovação e Dilemas Éticos na Inteligência Artificial – ITA). Sua pesquisa concentra-se em Ética na Inteligência Artificial, Filosofia Política e Filosofia da Tecnologia, com ênfase em contratualismo moderno (especialmente Hobbes), ética e IA, ética das máquinas, responsabilidade em sistemas autônomos, explicabilidade e a possibilidade de agência moral artificial. Sua pesquisa articula de modo integrado questões de ética, filosofia política, tecnologia e inovação. No campo da Ética na Inteligência Artificial, investiga problemas como a atribuição de responsabilidade em sistemas autônomos, a lacuna de responsabilidade, a explicabilidade algorítmica, a ética das máquinas e a possibilidade de agentes morais artificiais. Esses temas se conectam à Filosofia Política, especialmente ao contratualismo moderno e ao pensamento de Thomas Hobbes, explorando conceitos como soberania, representação, legitimidade, liberdade e racionalidade normativa. Ao mesmo tempo, sua investigação dialoga diretamente com a Filosofia da Tecnologia, analisando como a racionalidade tecnocientífica transforma práticas sociais, redistribui responsabilidades e engendra novos desafios éticos.

CV Lattes: <http://lattes.cnpq.br/7268737133400216>

FÁBIO LUIZ TEZINI CROCCO

É professor do Departamento de Humanidades (IEFH) do Instituto Tecnológico de Aeronáutica (ITA), membro do Laboratório de Cidadania e Tecnologias Sociais (LabCTS/ITA) e do Grupo de Pesquisa IDEIA – Inovação e Desafios Éticos da IA (CNPq). É doutor em Ciências Sociais pela Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP), na linha de pesquisa “Determinações do Mundo do Trabalho”, com estágio doutoral no Centro de Estudos Sociais (CES), vinculado à Universidade de Coimbra em Portugal (2014). Possui graduação em Ciências Sociais – Bacharelado (2005) e Licenciatura (2006) – e mestrado em Filosofia pela UNESP (2008). Atua com foco na Sociologia do Trabalho, nos estudos de Ciência, Tecnologia e Sociedade (CTS) e na Teoria Crítica.

CV Lattes: <http://lattes.cnpq.br/5950052772377300>

JOÃO DE FERNANDES TEIXEIRA

João de Fernandes Teixeira possui graduação em Filosofia pela Universidade de São Paulo (1977), mestrado em lógica e filosofia da ciência pela Universidade Estadual de Campinas (1982) e doutorado (PhD) em filosofia pela University of Essex (Inglaterra). Fez pós-doutorado nos Estados Unidos em 1998, no *Center for Cognitive Studies*, na *Tufts University*, sob a supervisão do Prof. Daniel Dennett. Participou do grupo de ciência cognitiva do Instituto de Estudos Avançados da USP. Foi professor na UNESP (campus-Marília) de 1982 a 1991. Em 1992 ingressou na Universidade Federal de São Carlos, na qual se aposentou como professor titular em 2016. Publicou mais de uma dezena de livros na área de filosofia da mente e ciência cognitiva, áreas nas quais é pioneiro no Brasil. Tem experiência na área de Filosofia, atuando principalmente nos seguintes temas: filosofia da mente, filosofia da psicologia e filosofia da neurociência e da inteligência artificial.

CV Lattes: <http://lattes.cnpq.br/8864985279295912>

JOHNNY CARDOSO MARQUES

Possui graduação em Engenharia de Computação pela Universidade do Estado do Rio de Janeiro/UERJ (2002). Mestre pelo programa de pós-graduação em Engenharia Aeronáutica e Mecânica na área de Sistemas Aeroespaciais e Mecatrônica pelo Instituto Tecnológico de Aeronáutica/ITA (2004). Doutor pelo programa de pós-graduação em Engenharia Eletrônica e Computação na área Informática pelo ITA (2016). Desde 2018, é Professor Adjunto da Divisão de Ciência de Computação e atualmente é chefe do departamento de Software e Sistemas de Informação. Orienta alunos de graduação e de pós-graduação do ITA na área de Engenharia de Software. Atuou entre 1998 e 2018 na indústria com atuação em computação e segurança (safety) para o setor aeronáutico com passagens pelas empresas Transbrasil (1998-2000), Varig (2000-2003) e Embraer (2003-2018). Foi Representante Credenciado em Engenharia de Software para Certificação Aeronáutica junto à Agência Nacional de Aviação Civil (ANAC) entre 2007 e 2018, atuando na aprovação de projetos como autoridade delegada. Pesquisa nas

áreas de engenharia de requisitos, qualidade de software para sistemas críticos, sistemas de informação, casos de segurança e ética e segurança com inteligência artificial. É autor de várias normas e publicações em engenharia de software na Associação Brasileira de Normas Técnicas (ABNT), Radio Technical Commission for Aeronautics (RTCA) e Institute of Electrical and Electronic Engineers (IEEE), como a RTCA DO-178C, RTCA DO-200B, ISO/IEC/NBR 62304 e ISO/IEEE 29148.

CV Lattes: <http://lattes.cnpq.br/6831514861193749>

JOSÉ ALEXANDRE MATELLI

Possui graduação em Engenharia Mecânica (1998), mestrado em Engenharia e Ciências Térmicas (2001) e doutorado em Engenharia e Ciências Térmicas (2008), todos da Universidade Federal de Santa Catarina. Atualmente, é professor livre-docente da Universidade Estadual Paulista (Unesp), no Departamento de Química e Energia da Faculdade de Engenharia e Ciências de Guaratinguetá. Atua também na Reitoria da Unesp, como assessor técnico da Coordenadoria de Desenvolvimento Profissional e Práticas Pedagógicas (CDeP3). Tem interesse e experiência na área de energia, atuando principalmente nos seguintes temas: análise e síntese de sistemas térmicos, projeto resiliente, cogeração, gás natural, células-combustível, inteligência artificial simbólica e engenharia de conhecimento.

CV Lattes: <http://lattes.cnpq.br/7690730276803550>

LIGIA PAVAN BAPTISTA

Doutora (2003), Mestre (1995) e Bacharel (1984) em Filosofia com ênfase em Ética e Filosofia Política, pela Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo. Especialista em Estudos da Paz e Resolução de Conflitos pela European University Center for Peace Studies (Áustria). Estágio de Pós-doutorado em Relações Internacionais/Universidade de Brasília. Consultora de organismos internacionais e instituições públicas do poder executivo federal. Responsável pela criação da Biblioteca Virtual sobre Corrupção, Biblioteca Digital do Ministério Público do Distrito Federal e Territórios e Biblioteca Virtual de Filosofia. Atua nas áreas de Filosofia Política, Teoria Política, Ciência Políti-

ca, Relações Internacionais, Ética, Ética na Administração Pública, Ética na Inteligência Artificial, Educação Superior. Direitos Humanos, Democracia e Cidadania, Filosofia e Gênero. Grupo de Pesquisa Ideia: Inovação e Desafios Éticos da IA (ITA/CNPq). Grupo de Pesquisa Direitos Humanos e Promoção da Igualdade de Gênero (PPGFIL/UnB). Sócia fundadora da Associação Brasileira de Relações Internacionais (ABRI) e do Fórum de Gestão do Ensino Superior nos Países e Regiões de Língua Portuguesa (FORGES). Secretária Regional da Sociedade Brasileira para o Progresso da Ciência (SBPC) no Distrito Federal (2015-2019). Membro do Comitê Consultivo da Rede de Mulheres Filósofas da América Latina (UNESCO). Editora do Boletim Chauí. Autora da obra “A Paz na Teoria na Política de Hobbes”. Professora de Ética e Filosofia Política da Universidade de Brasília e membro do Programa de Pós-Graduação em Filosofia – PPGFIL/UnB.

CV Lattes: <http://lattes.cnpq.br/1447623724695372>

MAURÍCIO CÉSAR DELAMARO

Possui graduação em Engenharia Mecânica pela Universidade Estadual Paulista – UNESP (1985), mestrado em Engenharia de Produção pela Universidade Federal do Rio de Janeiro (1989) e doutorado em Engenharia de Produção pela Universidade Federal do Rio de Janeiro (1997). Atualmente é professor assistente doutor da Universidade Estadual Paulista – UNESP. Tem como principais interesses a engenharia da sustentabilidade, desenvolvimento sustentável e educação para a sustentabilidade.

CV Lattes: <http://lattes.cnpq.br/3727732364475835>

RAFAEL EVANGELISTA

Rafael de Almeida Evangelista é graduado em Ciências Sociais (1998), mestre em Linguística (2005) e doutor em Antropologia Social (2010), pela Universidade Estadual de Campinas (Unicamp). É pesquisador do Laboratório de Estudos Avançados em Jornalismo (Labjor/Unicamp) desde 2003 e professor do Programa de Pós-Graduação em Divulgação Científica e Cultural da Unicamp desde 2011, tendo atuado também na coordenação do curso de mestrado. Possui espe-

cialização em Jornalismo Científico pela Unicamp e realizou pós-doutorado como pesquisador visitante no Surveillance Studies Centre da Queens University (Canadá), em 2018. Atua nas áreas de cultura da vigilância, capitalismo de vigilância (com ênfase no Sul Global), uso econômico de dados pessoais, trabalho informacional, tecnologias digitais, cibernética, cultura *hacker*, *software* livre e divulgação científica. Mais recentemente, suas pesquisas concentram-se em inteligência artificial, colonialismo digital, soberania digital, plataformas digitais e plataformação da educação. É vice-coordenador do Núcleo de Estudos da Criatividade (Nudecri/Unicamp), editor-chefe da revista ComCiência e membro da Rede Latino-Americana de Estudos em Vigilância, Tecnologia e Sociedade (Lavits). Atua no campo da governança da internet como representante eleito da comunidade científica no Comitê Gestor da Internet no Brasil (CGI.br) e conselheiro do Núcleo de Informação e Coordenação do Ponto BR (NIC.br). É autor do livro “Para além das máquinas de adorável graça: cultura hacker, cibernética e democracia”.

CV Lattes: <http://lattes.cnpq.br/4239876530989980>

RODOLFO AVELINO

Doutor pela Universidade Federal do ABC (2021); Mestre em TV Digital pela UNESP (2014). Possui graduação em Sistemas de Informação pelo Centro Universitário Fundação Santo André (2003). É Professor no curso de Engenharia e da Ciência da Computação do Insper. Membro eleito do Comitê Gestor da Internet no Brasil (CGI.br). Experiência como consultor e especialista em segurança da informação a mais de 29 anos. Pesquisador no Laboratório de Tecnologias Livres (LabLivre) da UFABC e componente da ONG Coletivo Digital. Possui experiência na área de Docência e Ciência da Computação, com ênfase em Redes de Computadores, Segurança da Informação e ambientes operacionais. Autor do livro “Colonialismo Digital: Tecnologias de rastreamento e a economia informacional” e um dos organizadores do livro “A sociedade de controle: manipulação e modulação nas redes digitais”. Faz parte do conselho da Artigo 19 (Brasil) e

da Rede Latino-Americana de Estudos sobre Vigilância, Tecnologia e Sociedade (LAVITS).

CV Lattes: <http://lattes.cnpq.br/0390038571471233>

WALTER TEIXEIRA LIMA JUNIOR

Pesquisador interdisciplinar em Tecnologia, Comunicação e Sistemas Cognitivos Artificiais. Professor de Ciência e Tecnologia do Instituto de Ciência e Tecnologia da Universidade Federal de São Paulo (Unifesp). Pós-doutorado em Simbiose Homem-computador no Departamento de Engenharia Mecatrônica da Escola Politécnica da Universidade de São Paulo (USP) e pós-doutorado em Sistemas Cognitivos e Comunicação na Universidade de Aveiro (POR). Atualmente é coordenador do Programa de Profissional Doutorado e Mestrado em Inovação Tecnológica da Universidade Federal de São Paulo (Unifesp). Bolsa de produtividade do Conselho Nacional de Desenvolvimento Científico e Tecnológico (2012-2021).

CV Lattes: <http://lattes.cnpq.br/5189390968642420>

1ª Edição - março de 2026



Diante da ascensão recente e popular da Inteligência Artificial Generativa (IA Generativa), como garantir que seu desenvolvimento e uso sejam éticos, justos e centrados no ser humano? Esta obra reúne debates e proposições essenciais, abordando desde a responsabilidade e transparência na IA, a formação docente e os desafios pedagógicos, até questões de direitos humanos, igualdade de gênero, crítica ao colonialismo e à colonialidade e defesa da soberania digital. Com base em perspectivas nacionais e internacionais, o livro propõe reflexões urgentes e caminhos para uma IA que amplie, e não substitua, o potencial humano.

