

**TEC** 28  
JUL./DEZ.  
2023

**COGS**

REVISTA DIGITAL DE  
TECNOLOGIAS COGNITIVAS

ISSN  
1984-3585

**Manual Ético para o uso  
da Inteligência Artificial  
Generativa**

Programa de Pós-Graduação em  
Tecnologias da Inteligência e Design Digital  
Pontifícia Universidade Católica de São Paulo

LE PENSEVR



LE PENSEVR  
BERT

## Expediente

TECCOGS – Revista Digital de Tecnologias Cognitivas, n. 28, jul./dez. 2023, ISSN: 1984-3585  
Programa de Pós-graduação em Tecnologias da Inteligência e Design Digital (TIDD),  
Pontifícia Universidade Católica de São Paulo (PUC-SP).



# PUC-SP

### DIRETOR CIENTÍFICO

Prof. Dr. Winfried Nöth  
PUC-SP

### VICE-DIRETORA CIENTÍFICA

Profa. Dra. Lucia Santaella  
PUC-SP

### EDITOR DO NÚMERO

Prof. Dra. Lucia Santaella  
PUC-SP

### EDITOR EXECUTIVO

Gustavo Rick Amaral  
UAM

### PROJETO GRÁFICO E DIAGRAMAÇÃO

Amanda Recke

### IMAGEM DA CAPA

Tomando como fonte uma foto da escultura de bronze de Auguste Rodin (Wikimedia Commons, File:Rodin-2014-06.jpg), a produção da capa é de autoria de Diego Franco a partir de prompts formulados utilizando as seguintes aplicações: CHAT GPT (Ajuste final do texto do prompt), MIDJOURNEY (Criação da imagem inicial), MAGNIFIC AI. Evolução da qualidade dos detalhes da imagem e PHOTOSHOP.

### CONSELHO EDITORIAL

Profa. Dra. Alexandra Primo  
UFRGS

Prof. Dr. André Lemos  
UFBA

Profa. Dra. Cláudia Giannetti  
Barcelona

Profa. Dra. Clarisse Sieckenius de Souza  
PUC-RIO

Profa. Dra. Diana Domingues  
UnB Gama

Profa. Dra. Geane Alzamora  
UFMG

Profa. Dra. Giselle Beiguelman  
USP

Prof. Dr. João Teixeira  
UFSCar

Profa. Dra. Luiza Alonso  
UnB

Profa. Dra. Maria Eunice Gonzales  
Unesp-Marília

Prof. Dr. Ricardo Ribeiro Gudwin  
Unicamp

Prof. Dr. Sidarta Ribeiro  
UFRN



**n. 28, jul./dez. 2023**

**Sumário**

**EDITORIAL** 5  
Lucia Santaella

**DOSSIÊ**

Por que é imprescindível um manual ético para a Inteligência Artificial Generativa? 7  
Lucia Santaella

**ENTREVISTA**

Uma visão multirreferencial da educação e tecnologia: uma entrevista concedida a Lucia Santaella 26  
Edméa Santos e Lucia Santaella

**ARTIGOS**

A inteligência artificial e os direitos fundamentais 45  
Gabriela Andrade e Anderson Röhe

Dilemas éticos da inteligência artificial para um ensino em metamorfose 56  
Luis Eduardo Viegas

Confusões e dilemas da antropomorfização das inteligências artificiais 67  
Anderson Röhe e Lucia Santaella

Implicações da inteligência artificial na educação 76  
Bruno Johnson Soares e Diego Franco

APREND.AI: Inteligências Artificiais Generativas no ensino superior 87  
Tina Fonseca e Lucila Campiglia

Guia ético para a Inteligência Artificial Generativa no Ensino Superior 108  
Diego Franco, Luís Eduardo Viegas e Anderson Röhe

**EXTRA DOSSIÊS**

Criatividade da Inteligência Artificial Generativa 119  
Ana Maria Di Grado Hessel e David de Oliveira Lemes

A periferia é protagonista: o uso ético e transformador da inteligência artificial 131  
Ana Maria dos Santos Rodrigues

**RESENHA**

Resenha de *Ética na inteligência artificial*, de Mark Coeckelbergh 151  
Dora Kaufman

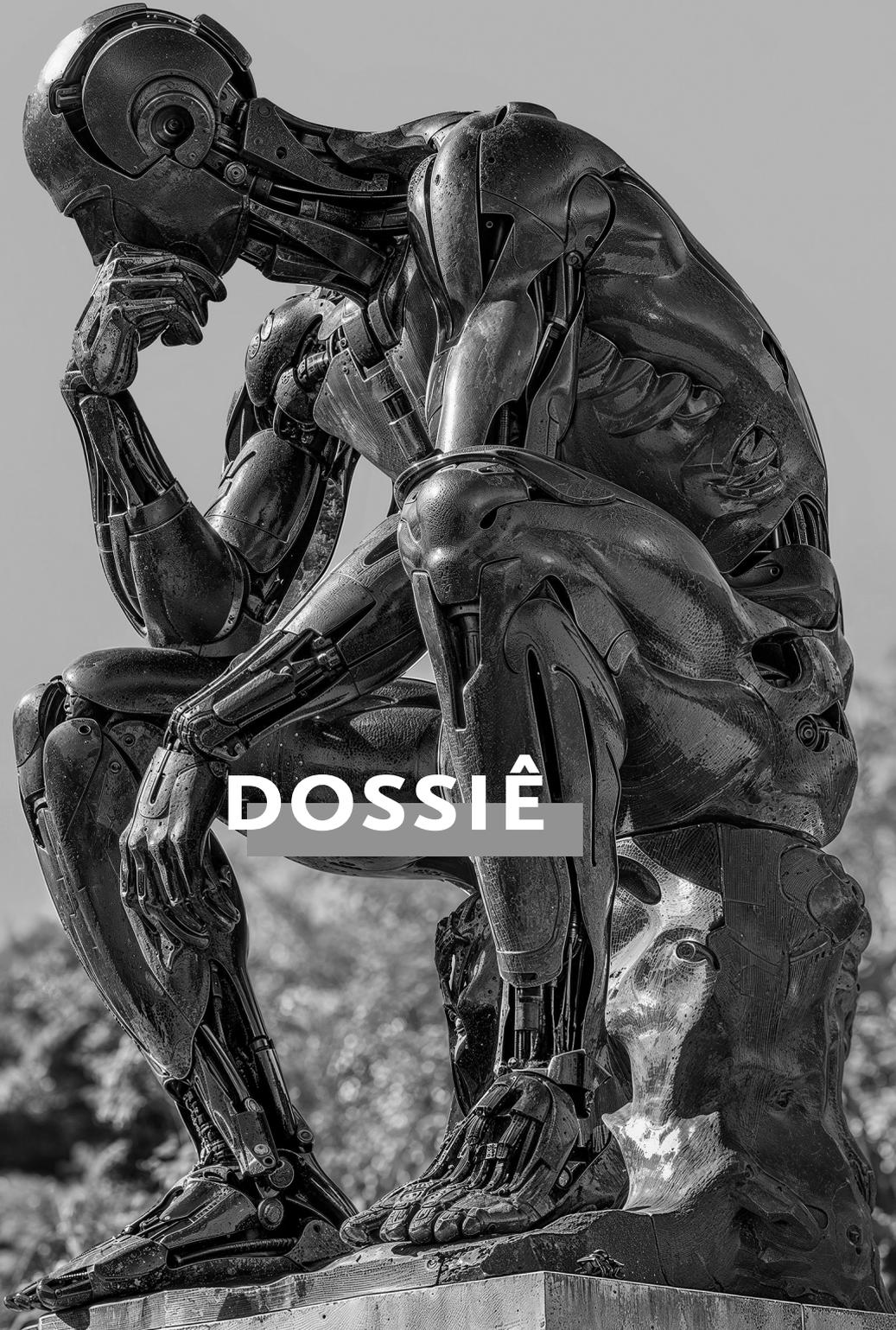
## Editorial

Por Lucia Santaella<sup>1</sup>

Supervisionado muito de perto por Lucia Santaella, este manual é fruto de um trabalho coletivo, o resultado de estudos e discussões desenvolvidas em vários meses até chegarmos a este ponto que o grupo julgou satisfatório. O trabalho nasceu da constatação de que alguma iniciativa deveria ser tomada em prol do desenvolvimento ético para o uso da Inteligência Artificial Generativa (IAG) no meio acadêmico. Esse uso não depende de qualquer conhecimento de qualquer espécie sobre quais são as determinações e como age a IAG. Desde a instauração das redes sociais, os usuários digitais acostumaram-se a ocupar seus espaços sem nenhum entrave. A IAG entrou nesse fluxo, em que tudo é facilitado para a habitação em uma ecologia tão amigável quanto ocultadora de todos os efeitos colaterais que provoca. Não é casual o verdadeiro tsunami provocado pelo ChatGPT, um multi tarefeiro conversador e solícito, sempre pronto não só a ajudar, mas também a fazer por nós. Aí mora o perigo. Este manual contém não apenas um guia ético, mas deve funcionar como um alerta ético no campo da educação. Na verdade, enquanto a IA preditiva lida sobretudo com riscos, a IAG tem na ética sua chamada magna. Educar para a ética tornou-se mandatório.

---

<sup>1</sup> É pesquisadora IA do CNPq, professora titular na pós-graduação em Comunicação e Semiótica e em Tecnologias da Inteligência e Design Digital (PUC-SP). Doutora em Teoria Literária pela PUC-SP e Livre-docente em Ciências da Comunicação pela USP. Publicou 56 livros e organizou 33, além da publicação de quase 500 artigos no Brasil e no exterior. Recebeu os prêmios Jabuti (2002, 2009, 2011, 2014), o prêmio Sérgio Motta (2005) e o prêmio Luiz Beltrão (2010). Orcid: <https://orcid.org/0000-0002-0681-6073>.



# DOSSIÈ

LE PENSEVR

# Por que é imprescindível um manual ético para a Inteligência Artificial Generativa?

Por Lucia Santaella<sup>1</sup>

**Resumo:** A tendência pervasiva da inteligência artificial, que já se anunciava há alguns anos, foi se intensificando cada vez mais, especialmente depois da emergência da Inteligência Artificial Generativa (IAG), em especial o Chat GPT, que foi colocado, pela start up Open AI, financiada pela Microsoft, para o uso público a partir do final de 2022. Protocolos de acesso muito facilitados, um quase nada, a partir de simples comandos verbais, tornam disponível um sistema capaz de atender às requisições informativas de um usuário, por meio de conversações sobre os mais variados assuntos, em uma linguagem que simula a capacidade humana de falar e de escrever. Distinta da Inteligência Artificial Preditiva (IAP), cujo desenvolvimento estava retido à competência dos especialistas, a IAG vem adquirindo as características de uma companheira disponível e solícita para a realização das mais variadas tarefas. Todas as áreas de produção humana relacionadas com linguagem estão sendo abaladas. Não haveria razões para que a educação ficasse de fora. Ao contrário. Entretanto, o Chat traz consigo um detalhe de suma importância: ele pode ser usado para fraudar ou para um uso honesto. Para atender a esse dilema ético, este manual, acompanhado de um guia foi elaborado.

**Palavras-chave:** manual ético; guia; Chat GPT; Inteligência Artificial Generativa

---

<sup>1</sup> É pesquisadora IA do CNPq, professora titular na pós-graduação em Comunicação e Semiótica e em Tecnologias da Inteligência e Design Digital (PUC-SP). Doutora em Teoria Literária pela PUC-SP e Livre-docente em Ciências da Comunicação pela USP. Publicou 56 livros e organizou 33, além da publicação de quase 500 artigos no Brasil e no exterior. Recebeu os prêmios Jabuti (2002, 2009, 2011, 2014), o prêmio Sergio Motta (2005) e o prêmio Luiz Beltrão (2010). Orcid: <https://orcid.org/0000-0002-0681-6073>.

## **Why an ethical manual is essential for Generative Artificial Intelligence?**

**Abstract:** The pervasive trend of artificial intelligence, which had already been announced a few years ago, intensified more and more, especially after the emergence of Generative Artificial Intelligence (IAG), in particular Chat GPT, which was made available by the start-up Open AI, financed by Microsoft, for public use from the end of 2022 on. Very easy protocols, almost nothing, from simple verbal commands, make accessible a system capable of meeting the user's informational requests, through conversations on the most varied subjects, in a language that simulates the human ability to speak and write. Distinct from Predictive Artificial Intelligence (IAP), whose development is left to the competence of specialists, IAG has been acquiring the characteristics of an available and helpful companion for carrying out the most varied tasks. All areas of human production related to language are being shaken. There would be no reason for education to be left out. On the contrary. However, Chat brings with it an extremely important detail: it can be used for fraud or for honest endeavors. To address this ethical dilemma, this Manual, accompanied by a guide, was created.

**Keywords:** manual of ethics; guide; Chat GPT; Generative Artificial Intelligenc

A partir de dezembro de 2022, o ChatGPT caiu no mundo feito um meteoro. A metáfora não é sem razão. No decorrer desse mês e naqueles que se seguiram, foi desmedido o volume de notícias, de colunas jornalísticas, de blogs e dos primeiros textos de natureza interpretativa. Até hoje o impacto não cessou, ao contrário, as tendências avaliativas são diversificadas, tornando o consenso cada vez mais perto do impossível, particularmente porque o Chat passou de 3.5 para a sofisticação do 4, prometendo o 5 e tornando-se inclusive apenas a ponta do iceberg de uma grande quantidade de sistemas em competição, além de outros subsidiários. Em suma: não é mais o mesmo aquilo que conhecíamos como revolução digital que, de resto, sempre foi uma revolução da ordem da progressão mesclada à dirupção. Nem a inteligência artificial (IA) é a mesma que ficou mais popularmente conhecida há uns quinze anos. Ela mudou e continua mudando. Quais as razões para tudo isso? Vamos por partes.

Embora o grande tema do momento, popularmente chamado de hype, seja de fato a IA, trata-se de uma área de investigação que teve seu início explícito em meados do século passado. Todos os textos que tratam do tema da IA com alguma seriedade, recorrem, mesmo que brevemente, ao seu histórico, uma recorrência importante para se evitar a displicência com o fato de que se trata de uma área de investigação científica que se desenvolveu com todos os rigores que são próprios da ciência. A IA nasceu com os mesmos propósitos que tem até hoje: desenvolver sistemas artificiais capazes de simular propriedades e habilidades que são próprias da inteligência humana. Seu ambiente de nascimento foi muito propício.

### **O nascimento da IA simbólica**

Conforme já desenvolvido em Santaella (2004, p. 73-92) e aqui retomado, quando as ciências da computação estavam emergindo, em 1956, foi realizada em Dortmund, USA, uma conferência com duração de seis semanas, que contou com a presença dos maiores especialistas em ciência da computação, tendo como tarefa estabelecer as bases de uma ciência da mente sob o modelo do computador digital. Foi dessa ideia de que o computador poderia ser um bom modelo para entender o funcionamento do cérebro humano que brotou a inteligência artificial cuja expansão interdisciplinar deu origem àquilo que passou a ser chamado de ciências cognitivas ou ciência cognitiva.

Segundo Teixeira (1998, p. 11-12), essa conferência rendeu frutos nos laboratórios de IA que foram fundados por John McCarthy e Marvin Minsky no MIT, Massachusetts, depois em Stanford, na Califórnia e tam-

bém em Pittsburgh, na Universidade Carnegie-Mellon, este sob iniciativa de Allan Newell e Herbert Simon (1972). Esses laboratórios criaram máquinas de jogar xadrez e outras máquinas inteligentes. O mais importante, contudo, era o que estava por trás da criação dessas máquinas, a saber, a busca das condições formais da atividade cognitiva, capazes de indicar o que é comum a todos os sistemas que exibem essa atividade, quer ela apareça em animais, máquinas ou humanos. Na época, o modelo computacional foi o escolhido devido à sua habilidade para simular processos cognitivos, com a possibilidade de se modelar a mente. Durante certo tempo, o “modelo computacional da mente”, isto é, o computador como metáfora da mente, constituiu-se como o paradigma clássico unificador das ciências cognitivas.

O modelo assentava-se pelo menos sobre dois pressupostos: (1) a relativa autonomia entre o *hardware* e o *software* das máquinas utilizadas para simular a inteligência, o que permite explicar o comportamento inteligente de qualquer sistema complexo sem pressupor o tipo físico ou biológico da inteligência de seus componentes; (2) a compreensão da mente como um conjunto de representações de tipo simbólico e regidas por um conjunto de regras sintáticas. Desse modo, o pensamento seria o resultado da ordenação mecânica de uma série de representações ou símbolos, ordenação esta que não pressupõe necessariamente a existência de um cérebro. Por isso, o aparato mental concebido como dispositivo lógico pode ser descrito por meio de um conjunto de computações abstratas. Simular a inteligência não implica a construção de máquinas com *hardware* específicos, mas sim o desenvolvimento de programas computacionais operando sobre dados ou representações. Chamado de IA simbólica, esse paradigma sedimentou-se a partir de final dos anos 1960, tanto nos trabalhos do grupo liderado por Newell e Simon quanto nos de Marvin Minsky e Seymour Papert (Teixeira, 1998, p. 36, 43).

O que está aí implicada é também uma mudança no conceito de inteligência que passa a ser definida como capacidade para produzir e manipular símbolos, tendo em vista a resolução de problemas. Em 1972, Newell e Simon haviam desenvolvido o conceito de “sistema físico de símbolos” para compreender como as pessoas resolvem problemas, uma vez que elas próprias são sistemas que manipulam símbolos. Mais tarde, em 1980, Newell reafirmou os fundamentos dos sistemas simbólicos físicos de modo mais sistemático. O conceito de sistema simbólico físico foi definido como “uma classe muito grande de sistemas capazes de produzir e manipular símbolos, sendo realizáveis dentro de nosso universo físico”. A hipótese é a de que esses símbolos, que são internos ao conceito

de sistema, são, “de fato, os mesmos símbolos que nós, seres humanos, produzimos e usamos todos os dias em nossas vidas”, o que significa que “os humanos são exemplos de sistemas simbólicos físicos, e, em virtude disso, a mente se insere no universo físico” (Newell, 1980, p. 136).

Conclusão, o pensamento era, então, visto como um sistema físico de símbolos, um tipo especial de máquina Turing que pode manipular símbolos. O símbolos são padrões físicos que podem ocorrer como elementos de um outro tipo de entidade que Newell e Simon chamaram de expressão (ou uma estrutura de símbolos), composta de um número de exemplares de símbolos que estão relacionados um depois do outro, de algum modo físico. Assim, os símbolos são como as letras de um alfabeto e as expressões como palavras e sentenças (Fetzer, 1991, p. 38). Depois de descrever o funcionamento de um sistema simbólico físico exemplar e depois de definir sua natureza essencial, Newell (1980, p. 172-173), considera o computador digital como um exemplo-chave para a realização de um sistema simbólico no nosso universo físico.

A teoria da mente que está por trás do paradigma computacional é o funcionalismo que exerceu domínio quase exclusivo nas ciências cognitivas até os anos 1980. Para o funcionalismo, “os estados mentais, tais como crenças e processos mentais, por exemplo, considerar e decidir, não são senão estados físicos descritos funcionalmente. O mesmo estado físico em sistemas diferentemente organizados pode levar a estados mentais diferentes; o mesmo estado mental pode ser realizado diferentemente em sistemas físicos diferentes”. O funcionalismo está baseado, portanto, na ideia de que a essência da natureza psicológica do estado ou processo mental não está na sua realização física particular, mas sim no seu papel computacional no sistema processador de informação (Garfield, 1987, p. 313-323). Assim, um mesmo papel funcional que caracteriza um determinado estado mental pode estar presente em sistemas nervosos completamente distintos. Esse é o chamado funcionalismo turinguiano para o qual a mente atualiza uma máquina de Turing no substrato biológico do cérebro.

Esse retrospecto dos primeiros tempos da IA é importante não só como documento histórico, mas também para lembrar que a IA simbólica não morreu, mas continua em novas formas convivendo com o sucesso retumbante alcançado pela atual IA baseada em redes neurais da aprendizagem profunda. Na verdade, a IA simbólica e a IA conexionista são hoje apenas duas entre outras três, a IA evolucionista, a IA bayesiana e a IA analógica, conhecidas como as cinco tribos da IA (ver Santaella, 2023, p. 37-43). Mas o sucesso atual alcançado pela IA conexionista não surgiu de um flash momentâneo, pois suas pesquisas tiveram início nos anos 1980.

## A emergência do conexionismo

Continuando com Santaella (2004, p. 73-82), a partir dos anos 1980, uma abordagem competitiva na concepção da mente começou a adquirir força nas ciências cognitivas. Trata-se do conexionismo que, sob o influxo de uma grande renovação das pesquisas sobre redes de neurônios formais, propôs a replicação da inteligência por meio da construção de redes neurais artificiais. Usando técnicas dotadas de propriedades que podem ser interpretadas em termos cognitivos, essas redes são capazes de aprender, reconhecer formas, memorizar por associações etc.

Já no movimento cibernético dos anos 1940, os cientistas estavam divididos entre duas concepções alternativas: estudar a mente ou o cérebro. Do lado do cérebro, estavam McCulloch e Pitts. Para eles, não se tratava de estudar o substrato físico do cérebro, mas as relações entre a lógica e o cérebro. Por isso mesmo, é nas pesquisas desenvolvidas por esses cientistas que o conexionismo encontra a sua paternidade, apesar de que essa origem não seja hoje devidamente lembrada. A falta de lembrança se deve, muito provavelmente, à grande repercussão que o modelo computacional da mente obteve nas ciências cognitivas durante décadas, o que deve ter provocado o esquecimento do bifurcamento das raízes que presidiu às suas origens.

Foram os trabalhos de G. E. Hinton, J. R. Anderson, D. E. Rumelhart e J. L. McClelland que, nos anos 1980, voltaram a chamar atenção para o estudo das redes neurais artificiais. Enquanto o modelo computacional da mente, também chamado de cognitivismo, estuda os processos mentais como computações abstratas, independentemente de suas formas específicas de concretização, o conexionismo pretende simular o cérebro como meio para emular a atividade mental. Os conexionistas tomam o cérebro humano como um dispositivo computacional em paralelo que opera com milhões de unidades similares aos neurônios.

Se computadores e cérebros têm como função principal processar informações, então redes neurais artificiais podem ser construídas para simular esse processo. As redes neurais constituem uma intrincada rede de conexões entre suas unidades que são dispostas em camadas hierarquicamente organizadas. Conectadas umas com as outras, unidades estimuladas via *inputs* externos excitam ou inibem outras unidades gerando padrões de conectividade. Diferentemente dos sistemas computacionais simbólicos, os conexionistas são sistemas dinâmicos compostos por um “conjunto de processos causais através dos quais as unidades se excitam

ou se inibem, sem empregar símbolos e tampouco regras para manipulá-los”. Pretende-se que esse conjunto de neurônios artificiais modele a cognição (Teixeira, 1998, p. 84).

Embora infinitamente menos complexo do que o cérebro, esse modelo da mente inspira-se na estrutura e modo de funcionamento do cérebro, chegando assim mais próximo da realidade biológica da mente. O que importa em sistemas desse tipo são os complexos padrões de atividade entre as múltiplas unidades que constituem uma rede. Para Teixeira (*ibid.*, p. 103), o conexionismo alinha-se com concepções filosóficas materialistas da mente, antecipando questões que viriam se tornar candentes com o desenvolvimento da neurociência cognitiva, da vida artificial e da robótica na década de 1990.

As pesquisas do conexionismo apontavam para futuros promissores, mas nos anos 1980 e 90 sua hora não havia ainda soado. Foi preciso esperar pela era do crescimento desmedido dos dados e pelo aumento da potência computacional para que elas pudessem passar a funcionar em simulações relativamente mais similares às operações do cérebro. Tendo como paradigma mestre o aprendizado, considerado como função magna do cérebro, o conexionismo encontrou o seu triunfo no aprendizado de máquina e no aprendizado profundo.

### **O triunfo do aprendizado de máquina e do aprendizado profundo**

Embora a aprendizagem de máquina e a sua sub-área de aprendizagem profunda sejam aquelas que mais recebem atenção, a IA constitui-se hoje em um campo hipercomplexo de investigações e de aplicações que apresenta como suas principais áreas, segundo Verma (2018, p. 5), os sistemas especialistas, instruções inteligentes auxiliadas por computador, processamento de linguagem natural, compreensão de fala, robótica e sistemas sensoriais, visão computacional e reconhecimento de cena, computação neural. Tudo isso, em rápido crescimento, está provocando um imenso impacto em vários campos da vida. As diversas técnicas aplicadas em IA são: “Rede Neural, Lógica Fuzzy, Computação Evolutiva, Instruções Auxiliadas por Computador e Inteligência Artificial Híbrida” (*ibid.*), entre outras.

Em uma definição abusivamente reduzida, inteligência significa “a capacidade de raciocinar, de desencadear novos pensamentos, de perceber e aprender” (*ibid.*, p. 6). Quando adicionado à palavra inteligência, o adjetivo “artificial” significa desenvolver computacionalmente máquinas capazes de funcionar à maneira da inteligência humana, tais como apren-

der (aquisição de informação e regras para usar as informações), raciocinar (usar as regras para chegar a conclusões aproximadas ou definitivas), autocorrigir-se e resolver problemas. Quando qualquer sistema se adapta às condições impostas por um ambiente, esse sistema é considerado inteligente. O que se extrai dessas considerações é o fato de que não é a condição psicologia humana e suas complexidades que a inteligência artificial busca simular, mas muito mais a capacidade de raciocinar e planejar para atingir determinados objetivos.

São muitas as tecnologias de IA. Conforme (Rouse, 2022), entre elas as mais conhecidas são:

- Automação como, por exemplo, a automação robótica de processos.
- Aprendizado de máquina, a ciência que faz um computador funcionar sem programação.
- Aprendizado profundo, subconjunto do aprendizado de máquina que, em termos muito simples, pode ser pensado como a automação da análise preditiva. Existem três tipos principais de algoritmos de aprendizado de máquina: 1. aprendizado supervisionado, no qual conjuntos de dados são rotulados para que padrões possam ser detectados e usados para rotular novos conjuntos de dados; 2. aprendizagem não supervisionada, na qual os conjuntos de dados não são rotulados e são classificados de acordo com semelhanças ou diferenças; e 3. aprendizagem por reforço, em que os conjuntos de dados não são rotulados, mas, após realizar uma ação ou várias ações, o sistema de IA recebe feedback.
- Visão computacional, concentrada no processamento de imagens baseado em máquina.
- Processamento de linguagem natural (PNL) é o processamento da linguagem humana por um programa de computador. Suas abordagens são baseadas no aprendizado de máquina e incluem tradução de texto, análise de sentimento e reconhecimento de fala.
- Robótica, um campo da engenharia focado no projeto e fabricação de robôs usados para realizar tarefas que são difíceis para os humanos realizarem ou executarem de forma consistente.

## Aplicações da IA

São também muitos os campos de incidência da IA. Os mais citados são:

- IA na saúde: o aprendizado de máquina está sendo usado para fazer diagnósticos melhores e mais rápidos do que os humanos e fornecer feedback médico básico.
- IA nos negócios: a automação robótica de processos está sendo aplicada a tarefas altamente repetitivas normalmente executadas por humanos. Algoritmos de aprendizado de máquina estão integrados em plataformas analíticas para descobrir informações para a melhoria dos atendimentos. Os chatbots foram incorporados aos sites para fornecer atendimento imediato aos clientes.

- IA na educação: usada para automatizar a avaliação, adaptando-se às suas necessidades e ajudando o aprendiz a trabalhar no seu próprio ritmo.
- IA nas finanças: aplicada a financiamentos por meio do recolhimentos de dados pessoais, inclusive fornecendo aconselhamento financeiro. O funcionamento geral dos sistemas bancários está impregnado de sistemas de IA.
- IA na lei: usada para automatizar análise de documentos e outras tarefas passíveis de serem programadas.
- IA na manufatura: incorporação de robôs no fluxo de trabalho humano (Rouse, 2022).

Entretanto, os setores de aplicações da IA são vastíssimos e tendem a se expandir. Um panorama das aplicações em vários setores e domínios da IA tão completo e atual quanto possível nos é fornecido por Duggal (2024, n.p.) ao qual recorreremos para complementar os quadros acima, agora com ênfase nas aplicações.

- Processamento de Linguagem Natural (PNL), usada para analisar e compreender a linguagem humana por meio de aplicativos como reconhecimento de fala, tradução automática, análise de sentimento e assistentes virtuais como Siri e Alexa.
- Análise de imagem e vídeo inclui a visão computacional para a análise e interpretação de imagens e vídeos aplicadas ao reconhecimento facial, detecção e rastreamento de objetos, moderação de conteúdo, imagens médicas e veículos autônomos.
- Sistemas de recomendação, baseados em IA são usados em comércio eletrônico, plataformas de streaming e mídias sociais para personalizar as experiências do usuário ao analisar as preferências, o comportamento e os dados históricos do usuário para sugerir produtos, filmes, músicas ou conteúdos relevantes.
- Assistentes Virtuais e Chatbots, com tecnologia de IA interagem com os usuários para lhes dar suporte e assistência personalizada ou executar tarefas.
- Jogos Algoritmos de IA são capazes de criar personagens virtuais realistas, tomando decisões inteligentes. A IA também pode otimizar gráficos de jogos, simulações físicas e testes de jogos.
- Casas Inteligentes e IoT são sistemas domésticos inteligentes para automatizar tarefas, controlar dispositivos e aprender com as preferências do usuário. A IA auxilia na funcionalidade e eficiência dos dispositivos e redes da Internet das Coisas (IoT).
- Ciber segurança diz respeito à detecção e prevenção de ameaças cibernéticas, ao analisar o tráfego de rede, identificando anomalias e prevenindo possíveis ataques.

As aplicações não param aí. Elas se expandem até se tornarem partes tão integrantes de nossa vida a ponto de se naturalizarem. Podem ser citadas como exemplo, o google maps, o waze, os filtros de realidade aumentada, Snapchat, ou “Lentes” que reconhecem características faciais, rastreiam movimentos e sobrepõem efeitos interativos nos rostos

dos usuários em tempo real. Bastante citados e discutidos são os carros autônomos e os dispositivos vestíveis rastreadores que monitoram e analisam dados de saúde. Por fim, o MuZero, desenvolvido pela DeepMind, alcançou um sucesso notável em jogos de tabuleiro complexos como xadrez, Go e shogi em um nível sobre-humano porque se auto-aprimora por meio do autojogo e do planejamento (Dugall, *ibid.*).

Todas essas aplicações são aqui citadas com o propósito de fornecer um panorama dos avanços da IA. Entretanto, existe um outro lado da IA que não pode deixar de ser lembrado. Trata-se das externalidades negativas da IA que também são muitas e preocupantes.

### As externalidades negativas da IA

Externalidade negativa é uma expressão proveniente da economia e ela significa a imposição de um custo a uma parte como efeito indireto de outra parte (Eldridge, 2024). A expressão passou a ser empregada com frequência no campo da IA para o qual o significado foi transposto com a finalidade de indicar os conflitos entre os limites da técnica e os direitos humanos (Kaufman *et al.*, 2023). Frequentemente mencionadas como externalidades negativas são, por exemplo, riscos aos direitos fundamentais, danos à democracia e ao meio-ambiente, reforço a discriminações não justificadas nas mais variadas esferas, coadjuvante em campanhas de desinformação e a intensificação do extrativismo, ou seja, da extração de recursos naturais. As externalidades negativas têm provocado uma necessária agitação em busca da regulamentação da IA que possa impedir ou minimizar riscos e efeitos colaterais. Segundo Bioni, Garrote e Guedes,

o contínuo surgimento de novas regulações direcionadas à AI, seja por meio de projetos de lei/regulamento ou de documentos internacionais de atores globais de relevância, revela a tendência global em que não se discute mais se, mas como se regular o uso desta tecnologia. Pela continuidade de produção de externalidades negativas de forma transversal, regulações setoriais não são suficientes. Isso, contudo, não afasta a necessidade de um arranjo de governança que navegue entre o geral e o específico justamente para traduzir normas de governança gerais às particularidades de um determinado contexto. Dito de outra forma, uma lei geral não exclui, mas, muito pelo contrário, abre espaço para que a regulação setorial floresça a partir de fundações comuns a diferentes setores da economia. (Bioni, Garrote e Guedes, 2023, p. 6)

O estudo realizado por Bioni *et al.* (2023, p. 611) levanta alguns tópicos capazes de auxiliar na organização do debate regulatório. São eles: (1) A regulação por meio da navegação entre o setorial e o geral. (2) A inovação responsável e resiliente socioeconomicamente. (3) Alvo regulatório plástico e uma regulação dinâmica e equilibrada (regulação assimétrica com base no risco). (4) Os vários modelos de regulação de risco. (5) Os

variados degraus da escada do risco. (6) O risco enquanto elemento dinâmico. (7) A difícil conciliação de uma regulação baseada em risco e em direitos – taxonomia de risco como um dos possíveis indicadores (proxy). (8) Avaliações de Impacto Algorítmico (AIA) públicas, inclusivas e sobre direitos sociais e não apenas individuais. (9) Uma regulação atenta aos aspectos sócio-técnicos-econômicos locais. (10) IAs Generativas (IAGs) e teste de stress das propostas de regulações de IA.

Como não poderia deixar de ser, dado o estado da arte da IA, os tópicos acima cessam na IAG, seguidos pela declaração de que, “o desafio das IAs Generativas é que, por se prestarem a diferentes finalidades (nem sempre previsíveis), tensionam o modelo regulatório baseado no risco, atualmente predominante no campo da IA, já que é inerentemente contextual” (Bioni *et al.*, p. 11). Entretanto, a hipótese que orientou o manual e o guia aqui presentes vão muito além de uma possível previsibilidade de riscos, pois a IAG não deve ser confundida com o que veio antes dela. Embora a IA precedente continue seu curso de investigações, aplicações e regulamentações, sem menosprezar a necessidade de regulamentação que também cabe à IAG, esta apresenta uma outra realidade, regida não só pelos riscos que apresenta, mas muito mais pelo uso ético que exige, cujas razões serão discutidas mais adiante, justificando a necessidade de um manual que aqui apresentamos, seguido de um guia para o uso da IAG.

## A IA preditiva e a IA generativa

Tanto a IAG apresenta uma realidade originalmente distinta que existem hoje dois títulos distintos de IA: a IA preditiva, de um lado, e IA generativa, de outro. Como funciona a IA preditiva e por que se chama preditiva? Imediatamente é possível inferir que ela é preditiva porque é capaz de prever resultados futuros. Como ela funciona para que isso seja possível?

Segundo Medha (2024), ela ingere grandes volumes de dados históricos de diferentes fontes, relevantes para o problema que lhe é colocado. Então os algoritmos de aprendizado de máquina analisam esses dados buscando tendências, padrões e relacionamentos entre variáveis.

Para isso, ela depende da modelagem estatística, ou seja, várias técnicas estatísticas e de aprendizado de máquina são usadas para, a partir dos dados, treinar modelos que sejam preditivos, ou seja, modelos que sejam treinados com o propósito de alcançar determinado resultado. Os métodos de modelagem mais utilizados são análise de regressão, árvores de decisão, redes neurais, previsão de séries temporais e modelagem de conjuntos.

Então, vem a fase da validação do modelo. Para isso, a exatidão e precisão dos modelos são não apenas rigorosamente testadas, quanto também os modelos são refinados até que o nível desejado de desempenho preditivo seja alcançado. A seguir, com os modelos razoavelmente precisos, passa-se para a simulação de cenário, quando diferentes cenários são simulados para o ajustamento dos parâmetros de entrada de modo a estimar previsões sob diversas condições. A etapa seguinte é a da implantação do modelo em ambientes de produção o que não impede que novos dados sejam continuamente inseridos nos modelos para gerar *insights* preditivos atualizados..

Por fim, vem a integração de processos dos *insights* preditivos “aos processos de negócios e fluxos de trabalho por meio de painéis, alertas APIs, etc., para permitir a tomada de decisões orientada por dados com base nas previsões do modelo”. Todo esse percurso torna a IA preditiva poderosa e valiosa para as corporações e organizações atuais, com o *surplus* de que os modelos tornam-se mais inteligentes com o tempo, à medida que processam mais informações (Medha, 2024, n.p.).

A IAG, por sua vez, diferencia-se da preditiva porque segue preceitos relativamente distintos, embora ainda se utilize de aprendizagem de máquina e redes neurais. Segundo Bharath (*apud* Lowton, 2023), a IA generativa e a IA preditiva representam paradigmas distintos no domínio da IA e do aprendizado de máquina, pois a generativa está voltada para a criação de conteúdo novo e original, como imagens, texto e outras mídias, aprendendo com os padrões de dados existentes. Por isso, pode-se dizer que auxilia em campos criativos e na resolução de novos problemas. A IA preditiva, por outro lado, usa padrões em dados históricos para prever resultados futuros ou classificar eventos futuros. Por isso, ajuda na tomada de decisões e na formulação de estratégias.

Isso não significa que sejam abordagens isoladas, pois podem se hibridizar em algumas situações. Assim, de acordo com Lawton (*ibid.*) a IAG pode ajudar a projetar recursos do produto, enquanto a IA preditiva pode prever a demanda do consumidor ou a resposta do mercado para esses recursos. Ainda também a IAG pode sintetizar dados realistas para aprimorar o conjunto de treinamento de um modelo preditivo, melhorando as capacidades preditivas. Enquanto a IA preditiva prevê eventos futuros analisando tendências históricas de dados para atribuir pesos de probabilidade aos modelos, a IA generativa cria novos dados, que podem estar na forma de texto e imagens.

Marcada a distinção entre esses dois paradigmas, o que importa ao manual que se segue é, antes de tudo, compreender qual é o modo de funcionamento da IAG que a capacita a produzir a variedade de resultados que produz, todos eles relacionados com a simulação de capacidades

semióticas humanas. A partir disso, pode-se compreender a explosão sociocultural e política provocada pela IAG e, diante disso, as razões que tornam imprescindível a existência de um manual ético para o seu uso na educação.

### **O que está por baixo do capô da IAG**

A IAG é um subconjunto do aprendizado profundo, mas de um tipo diferente, chamado de Modelo Gerativo que aprende com um conjunto subjacente de dados para gerar novos dados que imitam de perto os dados originais. Por meio do emprego de aprendizagem não supervisionada, esses modelos são usados principalmente para criar novos conteúdos, como imagens, texto ou até mesmo música, semelhantes àquilo que pode ser criado por humanos.

Os modelos generativos mais comuns são: Autoencodificadores Variacionais (VAEs), Redes Adversariais Gerativas (GANs), Máquinas Boltzmann Limitadas (RBMs) e Modelos de linguagem baseados em transformadores (*Transformers*). Os modelos geradores de textos estão baseados em grandes modelos de linguagem (*Large Language Models* – LLM) que é um tipo de aprendizagem de máquina treinado em um grande conjunto de dados de texto e que usa arquiteturas de redes neurais avançadas para gerar ou prever textos semelhantes aos humanos.

Os LLMs estão ligados ao processamento de linguagem natural (*Natural Language Processing* – NLP), que é um subconjunto da IA focado na interação entre computadores e humanos por meio da linguagem natural. Ele está habilitado a entender a língua humana e a se comunicar conosco na mesma língua (preferencialmente o inglês, vale lembrar), graças a algoritmos que ajudam os computadores a entender o contexto e o sentimento por trás das palavras e sentenças.

Por isso, os LLMs podem ser considerados como uma evolução dos modelos de processamento de linguagem natural. Embora estes últimos incluam uma ampla gama de modelos e técnicas para processar a linguagem humana, os grandes modelos se concentram na compreensão e geração de texto semelhante ao humano. Eles são especialmente projetados para prever a probabilidade de uma palavra ou frase com base nas palavras que a precedem, permitindo-lhes gerar textos coerentes e contextualmente relevantes. O processamento de linguagem natural utiliza uma ampla gama de técnicas, que vão desde métodos baseados em regras até aprendizado de máquina e abordagens de aprendizado profundo. O

LLM, por seu lado, usa principalmente técnicas de aprendizagem profunda para compreender padrões e contexto em dados de texto para prever a probabilidade da próxima palavra na sequência. Eles são um subconjunto da IAG que pode gerar muitos tipos de conteúdo como texto, imagem, vídeo, código, música etc., concentrando-se, portanto, na geração de texto.

A IAG, entretanto, é uma família mais ampla do que os sistemas voltados para a geração de texto. São conhecidos e bastante usados os sistemas de geração de imagem, como o Midjourney que usa um modelo de difusão, um tipo de modelo generativo que gradualmente adiciona ruído em uma imagem até que ela alcance o nível desejado de realismo. DALL-E usa uma rede neural treinada para produzir imagens a partir de comandos de textos. Foi anunciado com alarido um novo sistema, Sora, para geração de vídeos, muito mais potente do que eram capazes os sistemas anteriores.

Embora o ChatGPT (Transformador Gerativo Pré-treinado) tenha estado e continue estando na crista da onda desde dezembro de 2022, é bom saber que as aplicações dos LLMs não se reduzem a ele. São variados os setores e vale a pena conhecê-los. Gupta (2024) nos forcece uma lista:

- Assistentes Virtuais são os modelos LLM que analisam o comando humano e interpretam o seu significado, permitindo que os assistentes realizem ações requeridas pelos usuários.
- Tradução de idiomas são treinados em uma grande quantidade de dados de texto multilíngue, “o que lhes permite capturar distinções sutis, variações, contexto e complexidade de diferentes idiomas” (ibid.).
- Resumo: os modelos LLM podem resumir documentos ou artigos extensos, preservando as principais informações e pontos principais.
- Análise de sentimento com base em grandes quantidades de dados de texto, os LLMs conseguem compreender o contexto, as nuances e o tom da linguagem, identificando polaridades de sentimentos.
- Recomendações de conteúdo são realizadas pelas grandes plataformas a fim de fornecer aos usuários sugestões personalizadas e relevantes a eles, tendo como fonte a análise das interações dos usuários com conteúdos prévios.
- Bots de diálogo são aplicações para gerar texto coerente e contextualmente relevante semelhante ao humano, como reação a comandos dados pelos usuários. Dentre eles, o mais conhecido e usado é o ChatGPT cuja origem se reporta à introdução da arquitetura do transformador em 2017 que levou a OpenAI a lançar o GPT-1 em 2018. Vieram depois as versões GPT-2, GPT-3, GPT-3.5 e, então o GPT-4, um modelo pago, multimodal, o que significa que pode receber imagens e também texto como entrada. Já se fala em GPT 5.

### **Por que o ChatGPT 3.5 explodiu?**

É também educativa a razão pela qual o panorama histórico e situacional da IA, que foi elaborado acima, introduz este manual. É um modo de evitar uma tendência que vem tomando conta de muitos daqueles que se pronunciam sobre a IA caindo na mania do presentismo. Há poucos anos introduzi a ideia do presentismo como a reclusão no presente em si, um presente sem passado e sem futuro (Santaella, 2021, p. 121). Algum tempo depois encontrei essa mesma crítica tanto em Beck quanto em Crary. Para Beck (2018, p. 31), o mundo está sofrendo uma surpreendente metamorfose que exige a transformação do horizonte de referências e das coordenadas de ação, que são tomadas como constantes e imutáveis por posições que se mantêm aprisionadas no presentismo, ou seja, “posições que abolem o tempo para funcionar em tempo real, privilegiando o agora e nutrindo a ilusão da instantaneidade e da disponibilidade imediatas” (Crary, 2023, p. 85).

Infelizmente, na maior parte das apreciações e julgamentos socio-culturais que recebe, a IAG está sendo vítima de um presentismo agudo, como se tivesse surgido de um vácuo no tempo e no espaço. É certo que a IAG e mais particularmente o ChatGPT e congêneres produzem uma boa dose de inquietação e temores, pois com eles a IA chegou muito perto do humano, está roçando o cerne do humano devido à sua capacidade de dialogar como se fosse gente, penetrando, inclusive, em atividades concebidas como criativas em competição com aquilo que o humano considera seu tesouro e exclusividade magna. Diante disso, são até naturais as fantasias atemorizantes de que o humano está sendo usurpado de sua potencialidade mais preciosa. Contudo, só o conhecimento da história, do contexto mais amplo e, sobretudo, dos limites atuais da IAG pode evitar demonizações ou euforias despropositadas.

De outro lado, felizmente, o falatório acerca do fato de que a IA não é inteligente, baseado em noções muito antropocêntricas de inteligência, cessou até certo ponto. Ao mesmo tempo, intensificou-se o temor, por parte de alguns ex-desenvolvedores, hoje dissidentes, quanto ao desenvolvimento da IA geral ou forte, a saber, a IA que ultrapassaria o potencial humano até o ponto de levar nossa espécie a uma condição inócua ou mesmo ao extermínio. Embora não possamos descartar as incertezas e inseguranças quanto ao futuro, em nosso papel de pensadores, temos que olhar de frente para o modo como a IAG está penetrando em todas as nervuras de nossas vidas e para a maneira como crescem suas externalidades

negativas em relação aos direitos humanos, em especial, nesse caso, aos direitos autorais, perdas de empregos e outras consequências previsíveis e imprevisíveis.

Em nosso papel mais específico de educadores, papel em que especificamente este manual está contextualizado, olhar de frente significa reconhecer que a IAG está sendo explícita ou implicitamente usada pelos estudantes, sem que haja, pelo menos em nosso meio, guias claros que orientem os educadores em suas tarefas e tratamento minimamente seguro da questão.

Por que os estudantes estão inevitavelmente usando não é difícil de compreender, pois esse uso está ligado à razão primordial capaz de explicar por que a IAG explodiu com a força que demonstra, uma força que se localiza na grande e verdadeira diferença a se estabelecer entre a IA preditiva e a IAG. Estranhamente poucos têm reconhecido essa diferença, a saber: sem deixar de produzir seus efeitos em nossas vidas, cuja visibilidade mais imediata encontra-se nas semioses algorítmicamente guiadas dos nossos smartphones, a IA preditiva está reservada aos recintos dos desenvolvedores para a entrega às organizações e corporações. A IAG, por outro lado, caiu direto no colo de qualquer ser humano que disponha de um computador minimamente equipado, que saiba ler e escrever e que tenha disponibilidade para o diálogo, quando pode ver tarefas, que lhe custavam tempo, sendo realizadas rapidamente por um sistema solícito e prestativo.

A estratégia da Open AI foi magistral ao entregar, de modo gratuito e com facilidade ímpar de uso, o ChatGPT nas mãos dos usuários. Com isso, não apenas testou a receptividade dessa nova forma de IA quanto também se aproveitou dos comportamentos de uso para o aprimoramento do próprio sistema.

Depois de pouco mais de um ano o sistema ganhou aperfeiçoamento, sofisticou-se no GPT 4 e são poucos aqueles que não o estão experimentando. No campo da educação, proibir deve estar entre os piores caminhos, pois acaba por incentivar o uso oculto, cuja detecção, apesar de promessas, não é ainda segura. Ignorar significa alienar-se de um problema que exige atenção e encaminhamentos. A questão mandatória que se coloca é ética, o que envolve, antes de tudo, informar-se, conhecer, experimentar e avaliar para melhor agir. Este manual, seguido de um guia nasceu com esse propósito.

## Referências

BIONI, Bruno; GARROTE, Marina; GUEDES, Paula. *Temas centrais na regulação de IA: o local, o regional e o global na busca da interoperabilidade regulatória*. São Paulo: Associação Data Privacy Brasil de Pesquisa, 2023.

DUGGAL, Nikita. *What is artificial intelligence?* Simplilearn, 2 abr., 2024. Disponível em: <https://www.simplilearn.com/tutorials/artificial-intelligence-tutorial/what-is-artificial-intelligence>. Acesso em: 20 abr., 2024.

ELDRIDGE, Stephen. Negative externality. *Encyclopedia Britannica*, 15 mar., 2024. Disponível em: <https://www.britannica.com/topic/negative-externalitynegative-externality>. Acesso em: 17 mai., 2024.

FETZER, James. *Aspects of artificial intelligence*. Dordrecht: Kluwer, 1991.

GARFIELD, Jay L. (ed.). *Modularity in knowledge representation and natural language understanding*. Cambridge, MA: MIT Press, 1987.

GUPTA, Raja. Generative AI for beginners. Disponível em: <https://medium.com/@raja.gupta20/generative-ai-for-beginners-part-1-introduction-to-ai-eadb5a71fo7d>. Acesso em: 10 abr., 2024.

KAUFMAN, Dora; REIS, Patricia; JUNQUILHO, Tainá. Externalidades negativas da inteligência artificial: conflitos entre limites da técnica e direitos humanos. *Revista de Direitos e Garantias Fundamentais*, Sabta Lúcia. Vitória, v. 24, n. 3, p. 43-71, setembro/dezembro, 2023.

LAWTON, George. Generative AI vs. predictive AI: understanding the differences. TechTarget, 18 set. 2023. Disponível em: <https://www.techtarget.com/search/query?q=Generative-AI-vs-predictive-AI-Understandingthe-differences>. Acesso em: 2 abr., 2024.

MEDHA. Demystifying predictive AI: Definition and use. Fireflies.ai, 15 fev., 2024. Disponível em: <https://fireflies.ai/blog/predictive-ai>. Acesso em: 15 abr., 2024.

MERRITT, Rick. *What is a transformer model*. NVIDIA, 25 mar., 2022. Disponível em: <https://blogs.nvidia.com/blog/what-is-a-transformer-model/>. Acesso em 10 jun., 2023.

NEWELL, Allen. Physical symbol system. *Cognitive Science*, v. 4, p. 135-183, 1980.

NEWELL, Allen; SIMON, Herbert A. *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall, 1976.

ROUSE, Margaret. AI (Artificial Intelligence), 2022. Disponível em: <https://www.techtarget.com/searchenterpriseai/definition/AI-Artificial-Intelligence>. Acesso em: 10 dez., 2022.

SANTAELLA, Lucia. *Navegar no ciberespaço*. O perfil cognitivo do leitor imersivo. São Paulo: Paulus, 2004.

SANTAELLA, Lucia. Desafios e dilemas da ética na inteligência artificial. In: GUERRA FILHO, Willis S. *et al.* (org.). *Direito e Inteligência Artificial: fundamentos*, v. 1 – Inteligência Artificial, ética e direito. Rio de Janeiro: Lumens Juris, 2021. p. 109-136.

SANTAELLA, Lucia. *A inteligência artificial é inteligente?* São Paulo: Almedina, 2023.

TEIXEIRA, João de Fernandes. *Mentes e máquinas*. Uma introdução à ciência cognitiva. Porto Alegre: Artes Médicas, 1998.

VERMA, Mudit. *International Journal of Advanced Educational Research*, Delhi, v. 3, n. 1, p. 5-10, jan., 2018.



# ENTREVISTA

LE PENSEUR

# Uma visão multirreferencial da educação e tecnologia: uma entrevista concedida a Lucia Santaella

## A multi-referential and pluralistic view of education and technology: an interview with Lucia Santaella

Por Edméa Santos<sup>1</sup> e Lucia Santaella<sup>2</sup>

*Lucia Santaella (LS): Faça um breve retrospecto de sua atuação na área de educação e cibercultura...*

*Edméa Santos (ES):* Minha atuação profissional e acadêmica no campo da Educação e Cibercultura é diretamente implicada com minha história de vida e formação. Tudo começou com a minha autoria juvenil com a informática, quando ainda escrevíamos “Informática” com letra maiúscula. Comecei a programar com 15 anos de idade. Nos idos de 1987, ingressei num curso técnico de Processamento de Dados no ensino médio profissionalizante. Daí comecei a atuar como instrutora de Informática, ainda adolescente, mesmo antes de ingressar na universidade. Trabalhei ministrando aulas de informática em cursos de aperfeiçoamento, empresas e escolas, mas sempre na perspectiva tecnológica e profissional. Foram anos atuando com a docência, operação e programação de computadores de grande, médio e pequeno porte, bem como com a microinformática na era pré-interfaces gráficas.

Com a explosão do Windows e suas aplicações, fui deixando a área mais técnica e atuando, cada vez mais, com a docência no campo comercial. Ao ingressar no curso de ped-

---

<sup>1</sup> Professora titular-livre da Universidade Federal Rural do Rio de Janeiro. Bolsista Produtividade do CNPq. Cientista do Nosso Estado da FAPERJ. Professora do Programa de Pós-Graduação em Educação e Demandas Populares da UFRRJ. Membro do GT16 da ANPEd, membro do Conselho Diretor da ABCiber. Membro do Laboratório de Imagem da UERJ e do Observatório Formacce/UFBA. Editora-chefa da Revista Docência e Cibercultura e editora-gerente da Revista Interinstitucional Artes de Educar. Líder do GPODC – Grupo de Pesquisa Docência e Cibercultura. Site acervo: [www.edmeasantos.pro.br](http://www.edmeasantos.pro.br). Orcid: <http://orcid.org/0000-0003-4978-9818>.

<sup>2</sup> É pesquisadora 1A do CNPq, professora titular na pós-graduação em Comunicação e Semiótica e em Tecnologias da Inteligência e Design Digital (PUC-SP). Doutora em Teoria Literária pela PUC-SP e Livre-docente em Ciências da Comunicação pela USP. Publicou 56 livros e organizou 33, além da publicação de quase 500 artigos no Brasil e no exterior. Recebeu os prêmios Jabuti (2002, 2009, 2011, 2014), o prêmio Sergio Motta (2005) e o prêmio Luiz Beltrão (2010). Orcid: <https://orcid.org/0000-0002-0681-6073>.

agogia, comecei a articular o ensino da informática em contextos educacionais. Trabalhei anos em escolas da rede privada implementando laboratórios de informática, formando professores e atuando diretamente com currículos escolares articulando-os com o ensino de informática. Nesse contexto, já dispunha de referencial crítico e construtivista, o que já me destacava num cenário de maioria tecnicista. Com a chegada da internet e suas aplicações pré-web, tive oportunidade de interagir com sistemas de conversações online (BBS) e fui me implicando mais com a pesquisa propriamente dita.

Ingressei no Mestrado em Educação da Faced/UFBA em 1999 e passei a estudar a relação do Currículo com o Digital. Vi a internet tornar-se popular na era da WEB 1.0 e comecei a ter acesso às discussões específicas do campo da cibercultura. Sob orientação de Nelson Pretto, mergulhei nos estudos da cibercultura, estudando em língua portuguesa, em tempo real, a obra do filósofo e sociólogo Pierre Lévy, entre outros autores em ascensão no campo que começava a se forjar. Em 1998, antes mesmo de acessar o mestrado oficialmente, fiz no Rio de Janeiro um curso presencial com Pierre Lévy para discutir o livro *O que é o virtual?* Formação acadêmica sendo construída em meio à vivência interativa com a rede e suas interfaces com os territórios físicos e informacionais.

Simultaneamente vi a comunidade de científica da cibercultura nascer no Brasil. Costumava, como ouvinte, frequentar o GT Cibercultura da Compôs. Estudei profundamente obras dos colegas Nelson Pretto, André Lemos, Marco Silva, Marcos Palácios, Edvaldo Couto, Alex Primo, Eugênio Trivinho, entre outros e outras. Mas duas mulheres chamaram minha atenção com suas obras potentes. A primeira foi você, Lucia Santaella. Devorei e devoro tudo o que você produziu e produz. Você me ajuda cotidianamente a compreender os fenômenos da cibercultura, em suas mais plurais e dinâmicas fases, de forma interdisciplinar, o que muito me inspira a pensar, transpor e traduzir muitos dos princípios e fundamentos da comunicação na cibercultura para os estudos da educação na cibercultura, campo de que me tornei referência, junto e na parceira com meu grupo de pesquisa, o GPDOC – Grupo de Pesquisa Educação e Cibercultura.

A segunda mulher que muito me provocou à época foi a Diana Domingues, que me inspirou bastante com seu trabalho artístico. Os artistas fazem diferença na cultura e quando são cientistas, então, tudo fica ainda mais potente.

Não me esqueço jamais dos dois volumes, *A arte no século XXI e Arte, ciência e tecnologia*, em que Diana Domingues (2003; 2009) acionou uma potente comunidade científica internacional de artistas cientistas cibercul-

turais. As noções de hipertexto, simulação e interatividade, que nos são caras em termos teóricos e metodológicos, muito me influenciaram e me inspiraram do ponto de vista teórico e para a criação de meus dispositivos e atos de currículos, fazendo diferença no campo da educação na cibercultura.

Tempos depois, descobri que ela, Diana Domingues, tinha sido sua orientanda de mestrado e doutorado. Com o passar dos anos, outras mulheres foram tendo mais protagonismo no GT Cibercultura: Ivana Bentes, Fernanda Bruno, Raquel Recuero, Adriana Amaral, entre outras. Todas bastante eloquentes com suas pesquisas. Mas você, Lucia Santaella, sempre foi para mim uma inspiração, principalmente pela fartura, densidade e vitalidade de suas obras.

Durante a pesquisa de mestrado, estudei dois universos diferentes com uso de plataformas digitais na educação. A primeira foi “As árvores de Conhecimento”, a plataforma Gingo criada por Pierre Lévy e Michel Authier (1998). Realizei pesquisa de campo presencial durante o ano 2000 no Programa de Comunicação e Semiótica da PUC-SP, sob a supervisão do professor Rogério da Costa, que me recebeu em seu laboratório com muita generosidade. Por conta própria e sem bolsa de estudos, morei e fiz pesquisa em São Paulo por quase um ano. Foi uma experiência bastante ousada para uma jovem advinda das classes trabalhadoras do Nordeste brasileiro. Na capital paulista, no Lab do Programa, a conheci e só depois de um tempo me dei conta de quem era aquela professora, que ia linda e eloquente com seus orientandos criar e orientar projetos ciberculturais. Era você, Lucia Santaella. Não chegamos a interagir à época, mas tempos depois já estávamos parceiras em diferentes diálogos.

Durante o estágio na PUC, mergulhei no campo de uma intranet para estudar as conexões que o Gingo proporcionava com seus braços e brevês da Árvore de Conhecimento. A interface digital em rede que estudei, na verdade, era o que se convencionou chamar de AVA, ambiente virtual de aprendizagem, em que pude acompanhar as primeiras experiências de EAD *online* no Brasil.

Na sequência das pesquisas, produzi minha tese de doutorado pioneira na Faced/UFBA sobre educação online (Santos, 2005). Crie meu campo de pesquisa com a amiga e parceira Alexandra Okada, hoje pesquisadora da Open-UK. À época, Alexandra era doutoranda na PUC-Educação e juntas criamos um curso *online* para estudarmos e desenvolvermos nossas teses, agora sob a orientação de Roberto Sidney Macedo. Com esse trabalho defendi a tese/constructo da *Educação online como um fenômeno da cibercultura* (Santos, 2005), o que me fez conhecida dentro e fora do Brasil. Atualizei o

método da pesquisa-formação para a *Pesquisa-formação na cibercultura ou ciberpesquisa-formação* (Santos, 2015). Desenvolvi fundamentos e práticas de currículo e didática *online*. Minha obra sempre foi forjada com os estudos e práticas cotidianas ciberculturais.

Enfim, este trabalho vem sendo atualizado há 20 anos, se materializando em cada dissertação e tese que orientado junto ao GPDOC. Até a presente data, orientei 14 teses de doutorado, 17 dissertações de mestrado e 10 pós-doutoramentos, sem contar as monografias e TCCs de cursos de graduação e especializações. Tudo isso pode ser acessado no site [www.edmeasantos.pro.br](http://www.edmeasantos.pro.br). Costumo sugerir que acessem também o canal do GPDOC Rural no YouTube, em é possível acessar algumas defesas de tese e dissertações online: <https://www.youtube.com/@GPDOCRURAL/playlists>.

*LS: Você considera que educação e tecnologia é a mesma coisa que educação e cibercultura ou vê distinções entre ambas?*

*ES:* Não são a mesma coisa. O fato de se trabalhar com tecnologias digitais e educação não significa trabalhar com educação e cibercultura. Educação e cibercultura é quando operamos na interface educação e cultura contemporânea mediada pelo digital em rede, pelo viés dos estudos culturais específicos da era digital. Desde sempre me movimento na pesquisa, no ensino e na extensão universitária a partir de dois movimentos recursivos:

Primeiro: compreender a cena cibercultural, estudando seus fenômenos e eventos de forma interdisciplinar, multirreferencial, e com os cotidianos. Fazer imersão nas redes e nas suas interfaces entre os territórios físicos, informacionais e simbólicos para, com seus contextos, buscar inspirações para projetos de pesquisa e formação mais sintonizados com as dinâmicas comunicacionais, tecnológicas e pedagógicas do nosso tempo.

Segundo: desenvolver dispositivos, concebidos por mim como a inteligência pedagógica, mediada por tecnologias digitais em rede, que se materializam em atos e artefatos curriculares como dispositivos de pesquisa no contexto da cibercultura, sem dicotomizar as docências dos processos de investigação. Vem daí, criarmos e materializarmos educação e docência *online* na cibercultura, sempre a partir de projetos de ciberpesquisa-formação.

Muitos docentes e pesquisadores que operam com educação e tecnologia o fazem por outras bases teórico-metodológicas, a exemplo dos campos da tecnologia educacional (com desdobramentos para as áreas da EAD, informática na educação, mídia e educação, educomunicação, educação com tecnologias, entre outras), que é área no CNPq, infelizmente. Poderíamos ter uma área educação e cibercultura, pois já gozamos de maturidade enquanto campo e práticas de pesquisa. Pessoalmente trânsito e dialogo bem com diferentes comunidades e grupos de pesquisa nesta grande área da tecnologia educacional, inclusive no GT16 da ANPEd, do qual sou membro há 25 anos, já tendo atuado como coordenadora, parecerista *ad hoc* e parte do comitê científico.

Alguns colegas dessas áreas todas operam na lógica da aplicação e uso de tecnologia, inclusive a digital em rede, para implementar práticas educacionais já consolidadas em pedagogias instrucionais e ou forjadas em tecnologias analógicas e ou massivas. Assim, subutilizam o potencial tecnológico, comunicacional, pedagógico e epistemológico das redes e suas conexões. Exemplo clássico disso são os usos instrumentais e curriculares da EAD massiva, que passam longe das práticas interativas da educação *online*. No Brasil, temos uma infinidade de coletivos científicos que não se relacionam com educação e cibercultura, apesar de muitas vezes citarem nossos trabalhos e ou vampirizarem nossas referências. Nossa experiência consegue mapear erros grotescos e identificar apropriações equivocadas de nossas produções, e nesse ponto eu falo exatamente de nossos trabalhos, Lucia Santaella, tanto os seus na comunicação quanto os meus e de meus pares na educação. Até porque nem eu, nem você separamos essas grandes áreas da comunicação e educação da cultura, das ciências, das artes e das tecnologias.

Importante valorizar também os excelentes grupos de pesquisa e colegas que atuam como projetos incríveis nas mais diferentes áreas, a exemplo daqueles de base histórico-crítica, construtivistas, estudos culturais etc. Há trabalhos brilhantes em diferentes contextos de educação e tecnologias que não passam por inspirações específicas da educação e cibercultura. Em contrapartida, sempre identificamos gaps e problemas conceituais e metodológicos, uma vez que as especificidades do digital em rede não são valorizadas, sobretudo pela falta de pesquisa empírica e de quadros teóricos que deem conta das dinâmicas específicas do digital em rede. A internet não deve ser compreendida com referenciais da TV analógica, só para citar um exemplo mais simplório.

**LS:** *Você considera que educação e tecnologia é a mesma coisa que educação e cibercultura ou vê distinções entre ambas?*

**ES:** Não são a mesma coisa. O fato de se trabalhar com tecnologias digitais e educação não significa trabalhar com educação e cibercultura. Educação e cibercultura é quando operamos na interface educação e cultura contemporânea mediada pelo digital em rede, pelo viés dos estudos culturais específicos da era digital. Desde sempre me movimento na pesquisa, no ensino e na extensão universitária a partir de dois movimentos recursivos:

Primeiro: compreender a cena cibercultural, estudando seus fenômenos e eventos de forma interdisciplinar, multirreferencial, e com os cotidianos. Fazer imersão nas redes e nas suas interfaces entre os territórios físicos, informacionais e simbólicos para, com seus contextos, buscar inspirações para projetos de pesquisa e formação mais sintonizados com as dinâmicas comunicacionais, tecnológicas e pedagógicas do nosso tempo.

Segundo: desenvolver dispositivos, concebidos por mim como a inteligência pedagógica, mediada por tecnologias digitais em rede, que se materializam em atos e artefatos curriculares como dispositivos de pesquisa no contexto da cibercultura, sem dicotomizar as docências dos processos de investigação. Vem daí, criarmos e materializarmos educação e docência *online* na cibercultura, sempre a partir de projetos de ciberpesquisa-formação.

Muitos docentes e pesquisadores que operam com educação e tecnologia o fazem por outras bases teórico-metodológicas, a exemplo dos campos da tecnologia educacional (com desdobramentos para as áreas da EAD, informática na educação, mídia e educação, educomunicação, educação com tecnologias, entre outras), que é área no CNPq, infelizmente. Poderíamos ter uma área educação e cibercultura, pois já gozamos de maturidade enquanto campo e práticas de pesquisa. Pessoalmente trânsito e dialogo bem com diferentes comunidades e grupos de pesquisa nesta grande área da tecnologia educacional, inclusive no GT16 da ANPEd, do qual sou membro há 25 anos, já tendo atuado como coordenadora, parecerista *ad hoc* e parte do comitê científico.

Alguns colegas dessas áreas todas operam na lógica da aplicação e uso de tecnologia, inclusive a digital em rede, para implementar práticas educacionais já consolidadas em pedagogias instrucionais e ou forjadas

em tecnologias analógicas e ou massivas. Assim, subutilizam o potencial tecnológico, comunicacional, pedagógico e epistemológico das redes e suas conexões. Exemplo clássico disso são os usos instrumentais e curriculares da EAD massiva, que passam longe das práticas interativas da educação *online*. No Brasil, temos uma infinidade de coletivos científicos que não se relacionam com educação e cibercultura, apesar de muitas vezes citarem nossos trabalhos e ou vampirizarem nossas referências. Nossa experiência consegue mapear erros grotescos e identificar apropriações equivocadas de nossas produções, e nesse ponto eu falo exatamente de nossos trabalhos, Lucia Santaella, tanto os seus na comunicação quanto os meus e de meus pares na educação. Até porque nem eu, nem você separamos essas grandes áreas da comunicação e educação da cultura, das ciências, das artes e das tecnologias.

Importante valorizar também os excelentes grupos de pesquisa e colegas que atuam como projetos incríveis nas mais diferentes áreas, a exemplo daqueles de base histórico-crítica, construtivistas, estudos culturais etc. Há trabalhos brilhantes em diferentes contextos de educação e tecnologias que não passam por inspirações específicas da educação e cibercultura. Em contrapartida, sempre identificamos gaps e problemas conceituais e metodológicos, uma vez que as especificidades do digital em rede não são valorizadas, sobretudo pela falta de pesquisa empírica e de quadros teóricos que deem conta das dinâmicas específicas do digital em rede. A internet não deve ser compreendida com referenciais da TV analógica, só para citar um exemplo mais simplório.

*LS: Você é uma pesquisadora que não trabalha de modo solitário, mas atua em grupos colaborativos. Fale sobre isso?*

*ES:* Pesquisa em rede: Cada dissertação e tese que oriento é um subprojeto do projeto institucional por mim idealizado e desenvolvido, ligado aos programas de pós-graduação em que já atuei. Além disso, estabeleço uma rede muito intensa de conversas com colegas de dentro e fora do Brasil. Participo de projetos coletivos. Durante anos, estabeleci projetos com colegas do Nordeste, junto ao Programa Capes/Promob, só para citar um exemplo importante. Junto ao GT16 da ANPEd temos uma rede nacional bastante fértil com grupos de pesquisa de todo o Brasil que estudam na interface educação e comunicação.

Cocriamos números temáticos, livros, *ebook*, diferentes eventos, participamos de projetos e bancas. Incentivamos que nossos orientandos

interajam entre si e com coletivos de outros grupos também. Além das redes nacionais, também forjamos redes de pesquisa com colegas da Europa e Estados Unidos. Mais recentemente, estamos com redes também na América Latina e África. Mas, lamentavelmente, esse é um esforço muitas vezes pessoal e com poucos parceiros e redes de apoio. Durante minha carreira, realizei dois estudos de pós-doutorado, sendo que apenas um deles foi financiado com bolsa de estudos. Com mérito e com 18 anos atuando em mestrado e doutorado acadêmicos, sendo que durante 11 anos num Programa 7, obtive pouco sucesso com os financiamentos de agências. Durante anos, nossos estudos não eram compreendidos pelos pares da educação.

Só em 2022 fui contemplada com os editais do PQ/CNPq e CNE/Faperj, já percebendo sensível diferença para o desenvolvimento de projetos e difusão de seus resultados. Por outro lado, e por conta da minha capacidade de ser, viver e fazer redes, consegui ao longo dos anos estabelecer boas parcerias e conexões com colegas e grupos parceiros, principalmente para a divulgação de nossas pesquisas dentro e fora do Brasil.

Operamos com as dinâmicas da Ciência Aberta. Nossos trabalhos e rastros de autoria circulam em rede. Abrimos conteúdos e pesquisas desde sempre, uma vez que operamos na cibercultura. Ações que muitos grupos começaram a fazer um pouco antes ou durante a pandemia de covid-19, nossos grupos já faziam. Exemplos: *ebooks*, dossiês, revistas (criei e sou editora-chefa da revista *REDOC – Revista Docência e Cibercultura* (<https://www.e-publicacoes.uerj.br/re-doc>), divulgação científica em redes sociais, eventos científicos online (a exemplo do nosso e-DoC – Encontro Internacional de Educação na Cibercultura [<http://e-doc.pro.br/home>], que neste ano de 2024 já vai para sua quinta edição), transmissão de aulas públicas e partilhas de dispositivos de pesquisa e formação em diferentes espaços do ciberespaço. Estamos investindo em diferentes formas de comunicar a ciência que produzimos, o vídeo/filme de pesquisa é um dos recursos utilizados (Almeida; Martins; Santos, 2019).

São essas ações de nos forjam “rede”. Sem contar com a criação de oportunidades e projetos que visibilizaram trabalhos de grupo e colegas de todo Brasil, garantindo que estes também sejam realizados com parceiros internacionais. Pessoalmente, apresentei aos colegas da educação a nossa ABCIBER – Associação Brasileira de Pesquisa em Cibercultura, da qual sou membro fundadora e fui vice-diretora. O movimento também tem sido forte ao ensinar que colegas interajam também com nossas redes em Portugal, França e Inglaterra, por exemplo.

*LS: Quais as relações que você vê entre a educação e as redes sociais?*

*ES:* A relação é direta. Onde tem gente ensinando e aprendendo temos educação, seja esta formal, não formal ou informal. As redes sociais são *loci* em que seres humanos e objetos técnicos produzem e disputam narrativas e operam em rede de forma livre, aberta e caótica, revelando a sociedade que temos e seus valores éticos, estéticos e políticos, para o bem ou para o mal. Não há neutralidade nas redes sociais. Elas são para mim e o GPDOC um importante e fecundo campo de pesquisa. Nesses espaços costumamos ir ao que chamamos de “encontro com os etnométodos”, métodos próprios e jeitos de fazer na cibercultura, que subvertem práticas de comunicação massivas, até as lógicas desse capitalismo cognitivo e de vigilância que vivenciamos e que se agrava ainda mais com os fenômenos da datificação e plataformização da vida e da educação, sobretudo.

Por outro lado, encontramos nas redes sociais nossos mais cruéis dilemas societários. Há tempos, como você bem sabe e teorizou a respeito, acompanhamos a instalação de bolhas de pós-verdade nas redes sociais, o que nos desafia sobremaneira não só na pesquisa, mas, também e sobretudo, nas políticas públicas de acesso e acessibilidades na cibercultura. Nossas pesquisas buscam compreender as dinâmicas das redes sociais, como são forjadas e operadas pelos diferentes praticantes culturais, além de, como já falamos aqui, produzir dispositivos de pesquisa e formação nas e com as redes sociais. Pesquisamos nas e com as redes sociais já há alguns anos. Seguem alguns exemplos históricos: Amaral e Silva (2011) é uma etnografia na rede Facebook. A cibercultura e a educação em tempos de mobilidade e redes sociais: conversando com os cotidianos. Um livro organizado é *Facebook e educação: publicar, curtir, compartilhar* (Porto; Santos, 2014).

*LS: Como você vê a entrada da inteligência artificial no universo da educação?*

*ES:* Vejo com bons olhos: ela vai gerar oportunidades; mas também com preocupação. A cada entrada de desdobramentos das tecnologias digitais em rede, dilemas e etnométodos emergem. As tecnologias de IA generativa nos permitem operar de forma conversacional com as tecnologias e acervos semânticos. Requer formação crítica e multirreferencial, por parte dos usuários. Isso gera oportunidade de criarmos, mediarmos e avaliarmos diferentes práticas e atos intersemióticos de pesquisa e

formações. Intersemios e saberes científicos, digitais, urbanos, entre outros.

*LS: Qual sua experiência no campo da educação e da IA?*

*ES:* Estou há quase seis anos utilizando o GPDOC algumas interfaces de IA generativa, em especial para mobilizarmos saberes científicos, entre eles a capacidade de: Fazer boas perguntas em contexto; Analisar devolutivas, refutar e questionar as devolutivas da interface; Escrever e reescrever textos; Desenvolver argumentos; Traduzir textos em diferentes linguagens; Elaborar sínteses, quadros sinóticos, referências biográficas; Elaborar imagens a partir de textos e vice-versa; Elaborar planos de aulas e artefatos curriculares para nossos projetos de pesquisa-formação na cibercultura; Identificar notícias e falsas informações, refutando-as e criando processos contraditos.

Esses saberes são apenas alguns exemplos de mobilizadores para nossas práticas pedagógicas. Desenvolvemos dispositivos, oficinas, processos formativos. Por outro lado, questionamos os contextos nos quais as interfaces operam, buscando ampliar nossos repertórios sobre IA no contexto da plataformização da vida e da educação. Nossos trabalhos são sempre materializados em dispositivos. Sugerimos que conheçam alguns:

Link para nosso desenho didático Reglus, em autoria com Wallace Almeida (UFRRJ), que lança mão de IA para formar professores em contextos de checagem de notícias falsas. Usamos um *chat boot* (IA generativa) para produzir banco de notícias e combater as falsas. Vale ressaltar que tivemos a honra de ter você, Lucia Santaella, como membro titular na qualificação e defesa desse trabalho de tese. Você sabe que nossa autoria é aberta e permitimos reuso de nossas criações, contando com a ética dos usuários, porque atuamos na dinâmica da filosofia da abertura na cibercultura: <https://docenciaonline.pro.br/moodle/course/view.php?id=116>.

Link para capítulos e livros organizados sobre IA na Educação e aulas públicas: [www.edmeasantos.pro.br](http://www.edmeasantos.pro.br).

*LS: Qual a sua visão sobre a IA generativa de textos, o ChatGPT ou similares e de imagens nos processos de ensino e aprendizagem?*

**ES:** Trabalhamos exatamente com o universo das narrativas, imagens e sons em práticas de ciberpesquisa-formação. As narrativas textuais e imagéticas para nós são materiais de pesquisa, bem como forma e postura política de comunicar as pesquisas que forjamos. Narrativas para nós são posturas éticas, estéticas e políticas para democratizar o acesso à ciência e às pedagogias que criamos cotidianamente. As interfaces de IA generativa potencializam sobremaneira a nossa capacidade de produzir, refutar, criar e comunicar cientificamente. Na questão anterior, temos exemplos de alguns saberes que nos movem a criar, mediar e avaliar atos de currículos com uso dessas IAs generativas.

**LS:** *Quais os pontos positivos e os contrapontos negativos da IA generativa na educação?*

**Edméa Santos:** Para responder a essa questão, lancei mão do ChatGPT, IA generativa. Vejamos a nossa interação.

**Edméa Santos:** Quais os pontos positivos e os contrapontos negativos da IA generativa na educação?

**Chat GPT:** A IA generativa oferece uma série de pontos positivos e contrapontos negativos quando aplicada à educação. Aqui estão alguns deles:

**Pontos Positivos:**

1. **Personalização do Aprendizado:** A IA generativa pode criar conteúdo educacional personalizado, adaptando-se às necessidades individuais de cada aluno. Isso pode melhorar a eficácia da aprendizagem, permitindo que os alunos recebam materiais e atividades que correspondam ao seu nível de habilidade e estilo de aprendizado.
2. **Criação de Recursos Educativos:** A IA generativa pode ser usada para criar recursos educacionais, como tutoriais interativos, simuladores e materiais de aprendizado multimídia. Esses recursos podem tornar o aprendizado mais envolvente e acessível para os alunos.
3. **Feedback Automatizado:** A IA generativa pode fornecer feedback instantâneo aos alunos sobre seu desempenho em tarefas e atividades educacionais. Isso pode ajudar os alunos a entenderem seus pontos fortes e fracos e a melhorarem continuamente.
4. **Expansão do Acesso à Educação:** Com a ajuda da IA generativa, é possível criar sistemas de educação online e plataformas de aprendizado que oferecem acesso a recursos educacionais de alta qualidade para pessoas em todo o mundo, independentemente de sua localização geográfica ou situação socioeconômica.

### **Contrapontos Negativos:**

1. **Viés Algorítmico:** A IA generativa pode ser suscetível a viés algorítmico, o que significa que os sistemas podem reproduzir e amplificar preconceitos existentes presentes nos dados de treinamento. Isso pode resultar em desigualdades e injustiças na educação.
2. **Dependência Tecnológica:** A dependência excessiva da IA generativa na educação pode levar os educadores e os alunos a confiarem demais na tecnologia, em detrimento das habilidades humanas, como o pensamento crítico e a resolução de problemas.
3. **Privacidade e Segurança dos Dados:** A coleta e o armazenamento de dados pessoais dos alunos por sistemas de IA generativa podem levantar preocupações sobre privacidade e segurança. É importante garantir que os dados dos alunos sejam protegidos e utilizados de forma ética e responsável.
4. **Desumanização do Processo de Aprendizado:** O uso excessivo de IA generativa na educação pode levar à desumanização do processo de aprendizado, reduzindo a interação entre alunos e educadores e tornando a experiência de aprendizado menos personalizada e significativa.

Em resumo, a IA generativa oferece várias vantagens potenciais para a educação, mas também apresenta desafios e preocupações que precisam ser considerados e abordados cuidadosamente. É essencial encontrar um equilíbrio entre aproveitar os benefícios da tecnologia e garantir que ela seja utilizada de maneira ética, inclusiva e responsável.

**Edméa Santos:** Até que ponto a personalização do aprendizado é um ponto positivo? Autores sociointeracionistas já nos ensinaram que a aprendizagem colaborativa, em rede e entre pares, é muito mais potente que as práticas de personalização individualizadas.

**Chat GPT:** A personalização do aprendizado pode ser extremamente benéfica, mas é importante reconhecer que não é a única abordagem eficaz. A aprendizagem personalizada permite adaptar o conteúdo, o ritmo e as estratégias de ensino de acordo com as necessidades individuais de cada aluno. Isso pode aumentar a motivação, o engajamento e a eficácia da aprendizagem, pois os alunos podem avançar em seu próprio ritmo e se concentrar em áreas onde têm maior dificuldade.

No entanto, como você mencionou, as abordagens colaborativas, em rede e entre pares, também são extremamente valiosas. Elas oferecem oportunidades para os alunos aprenderem com os outros, desenvolverem habilidades sociais e colaborativas, e construir

conhecimento de forma coletiva. Além disso, a aprendizagem colaborativa pode promover a diversidade de perspectivas e a resolução de problemas em grupo.

Portanto, em vez de ver a personalização do aprendizado e a aprendizagem colaborativa como opostas, podemos reconhecer que ambas têm seu lugar e podem ser combinadas de maneira complementar. Por exemplo, os alunos podem se envolver em atividades colaborativas para desenvolver habilidades sociais e de resolução de problemas, enquanto ainda têm oportunidades para personalizar sua aprendizagem com base em suas necessidades individuais. Dessa forma, podemos aproveitar o melhor de ambos os mundos para criar experiências de aprendizagem ricas e eficazes.

**Edméa Santos:** E por que você me apresentou a personalização do aprendizado como ponto positivo?

**Chat GPT:** Apresentei a personalização do aprendizado como um ponto positivo porque, em muitos casos, ela pode ser uma abordagem eficaz para atender às necessidades individuais dos alunos. Aqui estão algumas razões pelas quais a personalização do aprendizado pode ser considerada positiva:

1. **Atende às necessidades individuais:** Cada aluno tem seu próprio estilo de aprendizagem, ritmo e nível de habilidade. A personalização permite adaptar o ensino para atender a essas diferenças individuais, garantindo que cada aluno receba o suporte de que precisa para ter sucesso.
2. **Aumenta a motivação:** Quando os alunos têm controle sobre sua própria aprendizagem e podem escolher atividades que são relevantes e interessantes para eles, eles tendem a ficar mais motivados e engajados no processo de aprendizagem.
3. **Melhora a eficácia:** Ao adaptar o conteúdo e as estratégias de ensino às necessidades individuais dos alunos, a personalização pode melhorar a eficácia da aprendizagem, garantindo que os alunos realmente compreendam e internalizem os conceitos.
4. **Fomenta a autonomia:** A personalização do aprendizado permite que os alunos assumam maior responsabilidade por sua própria aprendizagem, desenvolvendo habilidades de autorregulação e autonomia que são valiosas para o sucesso acadêmico e além.

No entanto, é importante reconhecer que a personalização do aprendizado não é a única abordagem eficaz e que há momentos em que as estratégias colaborativas e em grupo são mais apropriadas. O equilíbrio entre diferentes abordagens é essencial para oferecer uma educação abrangente e eficaz.

<b>Aspectos apresentados pelo ChatGPT</b>	<b>Meus comentários</b>
Personalização do aprendizado	Por mais que reconheça os estilos de aprendizagem de cada sujeito cognoscente e suas singularidades, não concordo com essa máxima. Sou a favor da criação e mediações de zonas de desenvolvimento proximais, e estas só são possíveis em rede, de forma colaborativa entre seres humanos com diferentes experiências junto com instrumentos culturais do nosso tempo. Não fazemos ZDP com aulas e práticas personalizadas, apenas por considerar que dentro de uma experiência mais colaborativa, possamos agregar atenções mais ou menos personalizadas.
Criação de recursos educativos	Considero esse item importante, mas é de fato potente quando os recursos educativos são criados com os estudantes e não necessariamente para os estudantes. Mais que recursos, temos de dar importância às mediações. Muitas teorias valorizam conteúdos e recursos e quase não dão importância ao trabalho docente de mediação didática e pedagógica.
Feedback automatizado	Feedbacks devem ser colaborativos, em rede e via grandes fóruns de conversas e significações coletivas. Automatizar feedback é valorizar a burocracia de respostas prontas e focadas em conteúdos. Pode promover práticas sem emoções e olhares complexos.

**Edméa Santos:** Reparem que nas bases de dados do ChatGPT os aspectos positivos não atendem a minhas concepções sobre educação na cibercultura. Vamos recuperar aqui os pontos e vamos dialogar com eles mais um pouco.

Sobre pontos negativos, vou concordar com a resposta do ChatGPT, pela pura falta de tempo para maiores desdobramentos. O que quis deixar como mensagem é que as IAs são sim boas fontes, mas requerem bons humanos em interlocução.

*LS: Quais mudanças se operaram na condição do professor quando seguimos a evolução das tecnologias desde que a cultura do computador se instalou?*

*ES:* As mudanças que operam na condição do professor quando seguimos a evolução das tecnologias, desde que a cultura do computador se instalou, na minha opinião, passa por uma mudança no paradigma comunicacional. Do docente unidirecional, forjado na lógica da transmissão, própria dos meios massivos e das telas que irradiam informação de polos exclusivamente emissores, para um docente dialógico, forjado na dinâmica da potência da interatividade, própria dos meios pós-massivos e das telas que pedem a intervenção física de um leitor interativo, imersivo, ubíquo e agora generativo como você, Lucia Santaella, já tão bem teorizou.

Compreender como operam cognitivamente e socialmente os leitores interativos, imersivos, ubíquos e agora generativos é tarefa urgente para o docente epistemologicamente curioso. Ir ao encontro de seus etnométodos é ponto de partida para qualquer docente em nosso tempo. Quem é meu estudante na escola básica ou na universidade? Como eles e elas operam no social na cibercultura e porque têm dificuldades de aprender e operar de forma autoral na educação formal? Quais as distâncias entre estes mundos, o mundo da cibercultura e os currículos ainda inspirados no século XIX? O que acontece com seus processos criativos, considerando suas autorias interseccionais, ou seja, suas implicações sociais e de classe, suas origens e ancestralidades étnico-raciais, suas identidades de gênero e sexualidades?

Responder a essas perguntas cotidianamente é um de nossos desafios, pois com as provocações seremos convidados a sair de nossas zonas de conforto, pois teremos de abrir nossos canais para os mundos de nossos estudantes, para com eles e elas cocriarmos atos de currículo situados e em contextos. Assim decolonizamos nossos repertórios históricos. Nos últimos anos estamos cada vez mais interseccionais, aprendendo com as mulheres no contexto feminista de quarta onda, com as feministas negras e pessoas trans. Estamos aprendendo com as periferias e comunidades tradicionais, com nossas salas de aula interativas, cartografando e etnografando nas redes e na relação cidade e

ciberespaço. A seguir vejamos alguns de nossos rastros: Atos de currículos híbridos: um projeto de aulas-oficinas no curso de Pedagogia da UFRRJ, ver Josiowicz e Almeida Velozo (2023).

Escrevivências ciberfeministas e ciberdocentes: narrativas de uma mulher durante a pandemia de covid-19: ver Santos (2022).

Etno(queer)hipergrafia, autorias trans no Instagram: ver Santos Coradini e Santos (2023).

Ciberfeminsimos e cibereducações: ver Santos, Fernandes e York (2022).

*LS: O que o professor ganha e perde no contexto atual da disseminação dos usos da IA generativa?*

ES: Para quem gosta de controle, o professor perde controle cada vez mais. Estudantes ganham autonomia para criar e plagiar e estudar menos, se sua autoria não for uma conquista de autonomia e honestidade intelectual. Ganha mais uma interface de apoio para suas práticas educativas e uma inesgotável fonte de informação e conhecimento sistematizado, cada vez mais autorais e multirreferenciais. Com a IA generativa, o professor ganha mais possibilidades de comunicar seus saberes. Estudar, preparar aulas nunca foi tão inventivo como outrora. Por outro lado, nunca foi tão arriscado para quem não estuda, pesquisa e ou tem a honestidade intelectual como fundante das práticas autorais.

As IAs generativas exigem um professor mais bem formado e autor de seus fazeres e saberes. Só com a formação forjada na pesquisa os professores poderão fazer bons usos das IAs generativas, para identificar erros, incoerências e até os chamados surtos informacionais. As IAs erram e erram muito, o que é facilmente identificável para quem domina e torna-se cada vez mais especializado em seu tema de estudos. Sou bastante citada na comunidade científica e tenho me deparado com artigos, dissertações e teses que apresentam referências que nunca foram minhas. Sempre que posso, vou sinalizando esses equívocos para estudantes e orientadores.

Precisamos dialogar bastante entre nós, com nossos grupos de pesquisa e parceiros com ética na pesquisa, direitos autorais e usos autônomos das IAs generativas. Além, de desenvolver cada vez mais nossas habilidades de diálogos densos, fecundos e corajosos sobre temas

tabus na academia, como, por exemplo: autorias vampirescas, roubos intelectuais, apropriação e plágio, entre outros.

Penso que nunca foi tão importante criarmos ambiências mais conversacionais, atividades que permitam que os alunos possam comunicar, inclusive oralmente suas descobertas, argumentações, resultados de pesquisa e apropriações teóricas, já que a escrita pode ser cada vez mais forjada com o protagonismo das IAs generativas. Por outro lado, a escrita em primeira pessoa poderá ser ainda mais valorizada, uma vez que ela expressa a autoria dos estudantes e docentes para além dos padrões objetivos e mecânicos produzidos pelas IAs generativas em termos textuais, mas também imagéticos. A sensibilidades das significações nunca nos foi tão cara em tempos de IA.

## Referências

ALMEIDA, Joelma; MARTINS, Vivian; SANTOS, Edméa. Videopesquisa e formação na cibercultura: atos de currículo e de pesquisa em educação. *Boletim GEPEN* [Grupo de Estudos e Pesquisas em Educação Matemática], Rio de Janeiro, n. 75, p. 140-153, jul./dez., 2019. Disponível em: <https://periodicos.ufrj.br/index.php/gepem/article/view/190>. Acesso em 10 fev. 2023.

AMARAL DA FONTOURA, Helena; SILVA, Marco (orgs.). *Práticas pedagógicas, linguagem e mídias: Desafios à pós-graduação em Educação em suas múltiplas dimensões*. Rio de Janeiro: ANPed Nacional, 2011.

DOMINGUES, Diana (org.). *A arte no século XXI: A humanização das tecnologias*. São Paulo: Editora UNESP, 2003.

DOMINGUES, Diana (org.). *Arte, ciência e tecnologia: Passado, presente e desafios*. São Paulo: Editora Unesp / Itaú cultural, 2009.

JOSIOWICZ, Alejandra J.; ALMEIDA VELOZO, Naira de (org.). *Humanidades digitais na América Latina: linguagens, metodologias e práticas de análise*. Campinas: Pontes Editores, 2023. Disponível em: [Humanidades digitais na América Latina: linguagens, metodologias e práticas de análise - Pontes Editores](#). Acesso em: abr., 2024.

LÉVY, Pierre; AUTHIER, Michel. *As árvores de conhecimentos*. Prefácio Michel SERRES. Trad. Monica M. Seineman. São Paulo: Escuta, 1995.

PORTO, Cristiane; SANTOS, Edméa Oliveira dos (orgs.). *Facebook e educação: publicar, curtir, compartilhar*. Campinas Grande: Editora da Universidade Estadual da Paraíba, 2014.

SANTOS, Edméa. *Educação online: cibercultura e pesquisa – formação na prática docente*. 2005. Tese (Doutorado) – Curso de Educação, Faculdade de Educação, Universidade Federal da Bahia, Salvador, Bahia, 2005.

SANTOS, Edméa. *Pesquisa-formação na cibercultura*. Santo Tirso: Whitebooks, 2015.

SANTOS, Edméa. *Escrevivências ciberfeministas e ciberdocentes: narrativas de uma mulher durante a pandemia Covid-19*. São Paulo: Pedro & João, 2022.

SANTOS, Edméa; FERNANDES, Terezinha; YORK, Sara Wagner. *Ciberfeminismos e cibereducações: Narrativas de mulheres durante a pandemia de Covid-19*. Salvador: Edufba, 2022.

SANTOS CORADINI, Fábio dos; SANTOS, Edméa. Etno(queer) hipergrafia: autorias trans no Instagram. *Revista da FAEEBA – Educação e Contemporaneidade*, v. 32, n. 72, p. 235–249, 2023.



**ARTIGOS**

LE PENSEVR

# A inteligência artificial e os direitos fundamentais

Gabriela Andrade<sup>1</sup>

Anderson Röhe<sup>2</sup>

**Resumo:** Certas aplicações da inteligência artificial têm o potencial ambivalente de afetar direitos fundamentais, justamente por ser capaz de gerar mais eficiência e celeridade nas atividades humanas. Mas a ambivalência provém principalmente das várias externalidades negativas da IA especialmente aquelas que ferem os direitos humanos. Este artigo se propõe a explicitar os riscos, em que contextos acontecem e quais iniciativas vêm sendo tomadas.

**Palavras-chave:** direitos fundamentais; inteligência artificial; normativas; risco

---

<sup>1</sup> Advogada graduada pela Universidade de São Paulo – USP. Mestranda em Tecnologias da Inteligência e Design Digital pela Pontifícia Universidade Católica de São Paulo. Associate Regulatory Counsel na Uber do Brasil Tecnologia Ltda. Lattes: <http://lattes.cnpq.br/1197255071956981>. Orcid: <https://orcid.org/0009-0001-5727-3347>.

<sup>2</sup> Advogado na Comissão Especial de Privacidade, Proteção de Dados e Inteligência Artificial da OAB-SP. Graduado pela UFF. Mestre em Políticas Internacionais pela PUC Rio. Doutorando em Tecnologias da Inteligência e Design Digital pela PUC SP. Orcid: <https://0000-0002-3104-6365>.

## **Artificial Intelligence and fundamental rights**

**Abstract:** Certain applications of artificial intelligence have the ambivalent potential of affecting fundamental rights, precisely because they are capable of generating more efficiency and speed in human activities. But the ambivalence comes mainly from the various negative externalities of AI in particular those that harm human rights. This article aims to explain the risks, in which contexts they occur and what initiatives are being taken.

**Keywords:** artificial intelligence; fundamental rights; regulations; risk

Ferramentas de Inteligência Artificial (IA) vêm modificando significativamente a forma com que seres humanos performam diversas atividades. O desenvolvimento e emprego da ciência e da tecnologia, entretanto, trazem em si uma natureza ambivalente, capaz de promover tanto benefícios quanto despertar novos temores. Enquanto as novidades trazidas pela IA, de uma forma geral, são capazes de gerar mais eficiência e celeridade nas atividades humanas e cadeias produtivas, há também uma série de riscos que já se manifestam ou que são previstos – alguns mais passíveis de concretização que outros.

Certas aplicações de IA têm o potencial de afetar direitos fundamentais (FRA, 2020, p. 5). Entre os exemplos, que podem colocá-los em risco, estão: algoritmos usados para recrutamento nos departamentos de recursos humanos que costumam preferir candidatos homens a mulheres; chatbots que, com pouco tempo de funcionamento, revelaram-se ‘racistas’; traduções automáticas que mostraram preconceito de gênero; e sistemas de reconhecimento facial que detectam melhor homens brancos, mas não mulheres negras (FRA, 2020, p. 15). Muitos destes riscos, portanto, provocam temor de que a IA possa ser usada de maneira prejudicial a indivíduos específicos, grupos sociais, ou à própria coletividade.

Algumas preocupações, aquelas menos passíveis de concretização, relacionam-se a um futuro hipotético no qual a IA melhorará a si mesma – e crescerá exponencialmente, mudando por completo a sociedade humana. Este ponto foi cunhado por Ray Kurzweil como "singularidade tecnológica", no livro *A singularidade está próxima: quando os humanos transcendem a biologia* (2018, p. 51). Mas enquanto tal singularidade não parece estar tão próxima quanto prognosticam as obras de ficção científica, visões céticas e cautelosas indicam que o avanço tecnológico precisa de endereçamento desde já. No livro *Desmistificando a inteligência artificial*, Dora Kaufman (2022) traz alguns blocos temáticos cuja associação à IA já demanda um cuidadoso debate:

- (i) O aumento do desemprego devido à automação: desde o início dos processos de industrialização, empregos que consistiam em atividades mais ou menos repetitivas passaram a ser substituídos por máquinas. Hoje, a IA permite que a substituição de uma pessoa por uma máquina possa ocorrer independentemente de um maior nível de repetitividade da atividade performada – o que gera, compreensivelmente, insegurança em relação a como manter o desenvolvimento tecnológico e, ao mesmo tempo, absorver essas pessoas no mercado de trabalho em outras funções, bem como evitar um aumento de desigualdades sociais.
- (ii) Justiça e ética no desenvolvimento e uso da IA, incluindo temas como fake news e vigilância, em especial quando é exercida por meio de tecnologias de reconhecimento facial.

(ii) Os vieses de tecnologias de IA que podem gerar discriminação a marcadores sociais específicos, tais como a dificuldade na obtenção de dados acurados de gênero e etnia para a formação de bases de dados mais próximas da realidade, acrescida de erros na aferição de dados demográficos de etnia por meio de sistemas de reconhecimento facial, entre outros.

(iv) O poder das big techs e o direito de propriedade na economia de dados.

(v) O uso da IA no setor da saúde, incluindo a criação de chatbots terapêuticos.

(vi) O impacto ambiental do uso da IA.

(vii) As modificações trazidas pela IA na cultura e nas interações humano-máquina.

### **Processo e prioridades da normatização da IA**

Enquanto muitos dos tópicos elencados acima relacionam-se entre si, guardadas as devidas proporções, o ponto comum entre todos revela que, além das preocupações estruturais iminentes, soma-se a elas o risco da potencialização de consequências negativas causadas pelo mau uso ou pelo uso mal-intencionado da IA. Isso deixa ainda mais claro que a problemática não é apenas técnica, mas também ética. Para enfrentar essa miríade de complexidades, que incluem tanto benefícios quanto riscos em potencial, a principal reação da comunidade tem ido, atualmente, na direção da elaboração e proposição de princípios regulatórios – guias de uso ético, princípios, valores, leis – que devem ser incorporados ao desenvolvimento e uso da IA, com o fim de aplacar os riscos da IA e, ao mesmo tempo, garantir benefícios e melhores resultados de seu uso.

Trata-se, porém, de uma tarefa tanto necessária e urgente, quanto inglória, considerando a velocidade com que a IA evolui e as diversas vertentes em que suas potencialidades ainda podem ser exploradas. Isso implica novos desafios para o sistema jurídico vigente, pois a IA segue avançando, muitas vezes de forma imperceptível, e, à medida que o faz, traz inovações e preocupações que, até o momento, não podem ser previstas com precisão. Significa dizer que a maior parte das iniciativas normativas correm o risco de se tornarem obsoletas em pouco tempo. Ainda assim, mantém-se a necessidade de princípios de orientação para o desenvolvimento e uso ético e responsável da IA. Na verdade, aponta-se para a necessidade da criação de ferramentas flexíveis e adaptáveis de regulamentação para tal.

Diante deste panorama, a quantidade de iniciativas normativas que propõe algum tipo de regulamentação da inteligência artificial é impressionante – e parece tentar acompanhar a própria velocidade de evolução da tecnologia. Em um levantamento feito por Joon Moon (2023) em As

*regras das leis para humanos, não humanos e transhumanos* (Código de Ur-Nammu, Três leis da Robótica e Princípios de Asilomar), tais iniciativas podem ser rastreadas desde 1942, a partir de *As três leis da robótica*, de Isaac Asimov. Assim, até o momento de conclusão desta pesquisa, os regramentos elencados, em ordem cronológica, foram:

- As três leis da robótica, Isaac Asimov (1942, EUA);
- Dez leis de Osamu Tezuka (1988, Japão);
- Declaração Mundial de Robôs Feira Internacional de Robôs de Fukuoka (2004, Japão);
- Carta de Ética para Robôs (2007, Coreia do Sul);
- As três leis da robótica responsável - do IEEE [Instituto de engenheiros elétricos e eletrônicos] (2009, Mundial);
- Cinco princípios éticos para a robótica do Conselho de Pesquisas em Engenharia e Ciências Físicas (EPSRC) e o Conselho de Pesquisa em Artes e Humanidades (AHRC, 2011, Reino Unido);
- Carta de Ética de Robôs (2012, Coreia do Sul);
- Oito leis da robótica de Shinpo Fumio – Keio University (2015, Japão);
- As lei de Satya Nadella (2016, EUA);
- Princípios de Parceria em IA (2016, EUA);
- Três princípios para criar IA segura (ou IA compatível com humanos) de Stuart Russell (2017, EUA);
- Três Regras para Sistemas de Inteligência Artificial - do diretor-executivo do Allen Institute for Artificial Intelligence (2017, EUA);
- Regras de Direito Civil sobre Robótica (2017, UE);
- 23 Princípios de Asilomar (2017, EUA);
- Os Princípios Gerais do Design Eticamente Alinhado – IEEE versão 2 (2017, Mundial);
- Princípios para transparência algorítmica e responsabilidade – USACM (2017, EUA);
- Diretrizes Éticas da Sociedade Japonesa para IA - da Japanese Society for Artificial Intelligence (JSAI) (2017, Japão);
- Princípios para a Governança da IA da The Future Society (2017, EUA e UE);
- Os dez principais princípios para a IA ética da UNI Global Union (2017, Mundial);
- Princípios da Política de IA do Instituto de Tecnológico de Informática (ITI) (2017, Mundial);
- Princípios de pesquisa e desenvolvimento (P&D) de IA - MIC (2017, Japão/G7);
- Princípios para a Era Cognitiva da IBM (2017, EUA);
- Desenvolvendo IA para Negócios com Cinco Princípios Fundamentais da editora Sage (2017, Reino Unido);
- Código IA da Câmara dos Lordes (2017, Reino Unido);
- Três ideias da Iniciativa de IA centrada no ser humano de Stanford Human-Centered AI (HAI) (2018, EUA);
- Princípios Harmoniosos de IA – HAIP (2018, China);
- Diretrizes Universais para IA da Coalizão The Public Voice (2018, Mundial);

- Rascunho dos Princípios de Utilização de IA do MIC (2018, Japão);
- Princípios éticos e pré-requisitos democráticos, Grupo Europeu de Ética em Ciência e Novas Tecnologias (EGE) (2018, UE);
- Carta aberta à Comissão Europeia de IA e Robótica (2018, Europa); OpenAI Charter da OpenAI (2018, EUA);
- Práticas gerais recomendadas para IA da Google (2018, EUA);
- Microsoft responsible AI principles (2018, EUA); (xxxiv)
- Princípios para Confiança e Transparência IBM (2018, EUA);
- Princípios orientadores da empresa SAP para inteligência artificial (2018, Alemanha);
- Diretrizes de Ética de IA do Grupo Sony (2018, Japão);
- Declaração de Montreal pelo desenvolvimento responsável da IA (2018, Canadá);
- Princípios de IA da Organização para a Cooperação e Desenvolvimento Econômico (OCDE) (2019, Mundial);
- Recomendações sobre a inclusão da África subsaariana na ética global da IA (2019, África) e
- Os oito princípios de ética da inteligência artificial (IA) da Austrália (2019, Austrália).

Embora a lista seja surpreendente, é pouco provável, pela velocidade de evolução da questão, que consigamos elaborar uma lista exaustiva de todas as iniciativas. Mas, à lista acima, podemos também incluir:

- *Recommendation on the Ethics of AI* (UNESCO, 2022);
- *Ethics in the Age of Disruptive Technologies: An Operational Roadmap*, lançado pelo *Institute for Technology, Ethics, and Culture* (ITEC), no Vaticano (Flahaux, Green e Skeet, 2023);
- As modificações mais recentes realizadas nos próprios corpos normativos já mencionados, como, por exemplo, modificações realizadas silenciosamente (Teo, 2023) pela *OpenAI* (2023) em sua lista de valores centrais para funcionários em outubro do mesmo ano.

Por outro lado, a quantidade de iniciativas normativas causa certa instabilidade, insegurança jurídica e perigo de sobreposições no ordenamento jurídico. Assim, por enquanto é possível indicar apenas pontos em comum entre a maior parte delas, já que as diferenças em metodologia, estrutura e valores protegidos podem criar diretrizes contraditórias. Dois pontos gerais, entretanto, surgem desde já como consequência óbvia da organização normativa de uma ferramenta tão ambivalente quanto a IA. Primeiramente, os princípios se voltam à proteção dos humanos para evitar consequências nefastas do uso da tecnologia. Mas também tentam tanto maximizar os benefícios de seu uso, quanto garantir que a tecnologia proporcionará melhoramentos e comodidade à condição humana.

## Princípios legais e éticos comuns

Tentando identificar princípios éticos comuns à maior parte das iniciativas, o Berkman Klein Center for Internet & Society elaborou o mapeamento *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI* (Fjeld, et al., 2023). O estudo partiu da análise de 36 documentos emitidos pela sociedade civil (5), governos (13), organizações intergovernamentais (3), setor privado (8) e governança multissetorial (7), e identificou os seguintes oito temas principais encontrados nas iniciativas normativas:

- Privacidade, presente em 97% dos documentos e incluindo consentimento (25%), controle sobre uso de dados (42%), habilidade de restringir o processamento de dados (11%), direito de retificação (8%), direito a esquecimento (6%), proteção de dados *by design* (17%), recomendação de leis de proteção de dados (17%) e outros princípios de privacidade (92%);
- Responsabilidade, presente em 97% dos documentos e incluindo verificabilidade e replicabilidade (36%), avaliações de impacto (53%), responsabilidade ambiental (17%); avaliação e fiscalização de resultados (47%), criação de um corpo de monitoramento (17%), capacidade de apelação (22%), remediação de decisões automatizadas (11%), responsabilidade civil/legal (31%); recomendações para adoção de nova regulações (53%), responsabilidade *per se* (69%);
- Segurança e proteção, presente em 81% dos documentos e incluindo segurança (61%), proteção (67%), proteção *by design* (8%), e predicabilidade (11%);
- Transparência e explicabilidade, presentes em 94% dos documentos e incluindo transparência (72%), explicabilidade (78%), dados em código aberto e algoritmos (28%), direito à informação (11%), notificação de decisões tomadas pela inteligência artificial em relação a indivíduos (19%), notificação em caso de interação com inteligência artificial (25%), e relatórios periódicos (17%);
- Justiça e não discriminação, presente em 100% dos documentos e incluindo não discriminação e prevenção de vieses (89%), representatividade e dados de alta qualidade (36%), justiça (56%), equidade (25%), inclusividade em impacto (42%), e inclusividade *by design* (47%);
- Controle da tecnologia por humanos, presente em 69% dos documentos e incluindo revisão humana de decisões automatizadas (33%), habilidade de auto-exclusão de decisões automatizadas (8%); e controle humano sobre a tecnologia (64%);
- Responsabilidade profissional, presente em 78% dos documentos e incluindo acurácia (19%), *design* responsável (44%), consideração de efeitos a longo prazo (33%), colaboração multissetorial (64%), e integridade científica (6%); e
- Promoção de valores humanos, presente em 69% dos documentos e incluindo valores humanos e prosperidade humana (44%), acesso à tecnologia (31%), e alavancamento para benefício da sociedade (64%).

O estudo também ressalta que apenas 64% dos documentos analisados trazem referências a direitos humanos – o que contrariou a expectativa da tendência de que iniciativas normativas relacionadas à IA passassem a focar cada vez mais seus esforços na promoção e preservação de

tais direitos. Uma hipótese levantada pelo estudo para tal discrepância é a de que possivelmente as entidades responsáveis pela elaboração dos respectivos documentos tivessem um foco voltado a leis consideradas como mais relevantes ao assunto, como as leis de proteção de dados.

Acompanhamos esse estudo com uma segunda hipótese que partiu de uma observação também levantada pelo trabalho de Joon Moon (2023): grande parte das iniciativas normativas têm considerado a IA como agente de uma ação, que deve ou ser coibida ou cumprida, dependendo do caso. Sendo assim, por associação lógica e semântica, faz sentido se pensar em uma diminuição de menções a direitos humanos – uma vez que o protagonismo dos textos normativos não está necessariamente no humano, mas sim na IA.

### **A antropomorfização ética e legal da IA**

A atribuição de obrigações éticas ou jurídicas (e de direitos) aos não humanos não é uma novidade, mas não significa ser algo menos complexo. Vale o exemplo no campo da ética e dos direitos dos animais, em que se discute se direitos são atribuídos a animais por sua natureza senciente, ou se são atribuídos a animais como um prolongamento dos direitos humanos.

Pode-se dizer que a mesma discussão se estende à atribuição de obrigações a criaturas não humanas – caso em que se atribuem obrigações à IA – mas na direção contrária. Afinal, obrigações são atribuídas à IA porque se admite que sua inteligência a torna parte da comunidade moral humana, ou como um prolongamento das obrigações atribuídas aos humanos? Ou seja: entende-se que a IA possui ingerência suficiente para o cumprimento de obrigações, ou se entende que as obrigações éticas fundamentais ao desenvolvimento e uso da IA estão na comunidade humana que os cercam?

Não se trata de negar a inteligência às máquinas. Pode-se considerar que a inteligência é um atributo inerente também à IA (Santaella, 2023a), como seu próprio nome diz. Não se trata, também, de ignorar que a IA ainda está em evolução, não sendo possível afirmar ou negar que a inteligência das máquinas se desenvolverá de formas imprevistas – o que poderia levar a enquadrá-la de diferenciadas formas na comunidade social humana, atribuindo-lhe distintas funções e obrigações. Por fim, igualmente não se trata de discutir a adequação dos conceitos de “agência” e “responsabilidade” para o endereçamento de discussões voltadas à IA. Apesar de serem todas discussões relevantes ao tema, são, também,

tangenciais. Trata-se, sim, de estabelecer se a presença de inteligência, como a temos hoje, seria suficiente para que obrigações éticas possam ser atribuídas à IA. E, mais ainda, se a criação de iniciativas normativas que partem desse pressuposto é, de fato, eficaz.

Enquanto não se propõe um posicionamento específico sobre o tópico, entendemos que discussões mais focadas da questão exigirão recortes temáticos, tanto no sentido das tecnologias específicas que estão sendo discutidas, quanto em relação ao campo/setor de aplicabilidade delas. Isso dito, e tentando evitar o desenvolvimento de um manual ético que apenas ecoe a vasta produção de iniciativas normativas que revisamos anteriormente, propõe-se aqui a criação de um manual ético voltado especificamente ao uso de ChatGPT e de IA Generativa (IAG) de imagens sintéticas no campo do ensino superior. Considerados, assim, os potenciais usos dessas tecnologias no ambiente acadêmico, de saída, não parece fazer sentido atribuir obrigações à IA por si só, mas sim criar diretrizes para seu desenvolvimento e usos, por parte de humanos, neste campo específico.

### **Recorte temático: ChatGPT e IAG de imagens no ensino superior**

Um dos campos com maior potencial para o uso e conseqüente desenvolvimento da IA é o ensino superior. O advento da IAG cria uma situação sem precedentes em que não são claras as orientações para o seu uso ético, produtivo, eficaz e justo no ambiente acadêmico. Vale lembrar que uma semana após o lançamento do Chat GPT, em novembro de 2022, a ferramenta já possuía mais de um milhão de usuários. A interação com o GPT dá aos usuários a impressão de dialogar com um ser pensante e simpático. Em *Há como deter a invasão do ChatGPT?*, Santaella (2023b) o descreve como um "tarefeiro multitarefas" com imensas possibilidades de geração de conteúdo. E explica que tais habilidades se dão pelo fato de que os algoritmos que aprendem as associações estatísticas entre palavras e frases foram criados para executar quase tudo de que dispõe a linguagem verbal humana e os processos cognitivos concernentes à linguagem verbal e não verbal. Sheenan (2023) indica três fatores que causaram um acelerado uso da IAG Generativa na educação:

- (i) Amplo acesso à baixo (ou nenhum) custo;
- (ii) Engajamento via interfaces de usuário baseadas em texto e/ou imagem – o que acelera a geração de resultados em formato escrito, visual ou em código.
- (iii) A percepção de qualidade e a escala do treinamento da ferramenta em amplos modelos de linguagem, proporcionando resultados críveis.

O autor (*ibid.*) aponta para alguns usos potenciais dessas tecnologias no ensino superior, tais como:

- (i) uso, por parte de estudantes, para pesquisa, desenvolvimento de conteúdo e tarefas acadêmicas;
- (ii) uso, por parte da equipe administrativa, para a elaboração de relatórios, análise de dados e melhoria do suporte a alunos, e
- (iii) capacidade de uso para o planejamento de aulas e desenvolvimento de material para professores.

A princípio parece bem direcionado, embora suscinto, deixando muitas dificuldades não abordadas. Isto porque não se deve ignorar o que foi levantado anteriormente: a IAG é uma ferramenta ambivalente. Ela traz dilemas inerentes a todas as suas potencialidades. As bases de dados podem não estar devidamente atualizadas, além de sujeitas a vieses dos humanos que as constituíram. As respostas oferecidas não necessariamente são corretas, e não podem ser tomadas como verdadeiras sem conferência. As tecnologias nem sempre são capazes de se esquivar do mal uso por seu usuário, nem de dar respostas adequadas a pedidos inapropriados, uma vez que a IA não está imbuída de senso comum.

Fica perceptível, portanto, a relevância do fator humano no estabelecimento dos formatos éticos, apropriados e eficazes de tais ferramentas. Não se trata de uma resposta propriamente dita, mas resume a pergunta central que será endereçada por este manual: como aplicar a IAG no ambiente do ensino superior, aproveitando suas potências e mitigando seus dilemas? Uma questão que merece, de fato, um guia orientador, mas devidamente antecipado por discussões que sejam capazes de reconstruir o seu contexto.

## Referências

- ASILOMAR AI PRINCIPLES, jan. 2017. *The Future of Life Institute*. Disponível em: <https://futureoflife.org/ai-principles/>. Acesso em mai, 2024.
- ASIMOV, Isaac. In: ASIMOV, I. *Runaround*. New York: Spectra Books, 1942.
- FJELD, Jessica, *et al.* Principled artificial intelligence: mapping consensus in ethical and rights-based approaches to principles for AI. *Berkman Klein Center for Internet & Society*, 2020. Disponível em: <https://dash.harvard.edu/handle/1/42160420>. Acesso em: 14 fev. 2024.

FLAHAUX, José Roger; GREEN, Brian Patrick; SKEET, Ann. Ethics in the age of disruptive technologies: an operational roadmap – the ITEC handbook. *The Institute for Technology, Ethics, and Culture and the Markkula Center for Applied Ethics*, Santa Clara: Santa Clara University, 2023. Disponível em: <https://www.scu.edu/institute-for-technology-ethics-and-culture/itec-handbook/>. Acesso em: 18 abr. 2024.

FRA (EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS). Getting the Future Right: Artificial intelligence and fundamental rights. Luxembourg: *Publications Office of the European Union*, 2020. Disponível em: [https://fra.europa.eu/sites/default/files/fra\\_uploads/fra-2020-artificial-intelligence\\_en.pdf](https://fra.europa.eu/sites/default/files/fra_uploads/fra-2020-artificial-intelligence_en.pdf). Acesso em: 15 fev. 2024.

KAUFMAN, Dora. *Desmistificando a inteligência artificial*. Belo Horizonte: Autêntica, 2022.

KURZWEIL, Ray. *A Singularidade está próxima: quando os humanos transcendem a biologia*. Tradução Ana Goldberger. São Paulo: Itaú Cultural. Iluminuras, 2018.

MOON, Joon. As regras das leis para humanos, não humanos e transhumanos (Código de Ur-Nammu, Três leis da Robótica e Princípios de Asilomar). Dissertação de Mestrado. Repositório de Teses e Dissertações Defendidas do Departamento de Filosofia da USP, 2023. Disponível em: <https://filosofia.fflch.usp.br/node/10001>. Acesso em: 15 fev. 2024.

OPENAI. Join us in shaping the future of technology. *OpenAI*, 2023. Disponível em: <https://openai.com/careers>. Acesso em: 15 fev. 2024.

SANTAELLA, Lucia. *Neo-humano: a sétima revolução cognitiva do sapiens (Comunicação)*. São Paulo: Paulus Editora, 2022.

SANTAELLA, Lucia. *A inteligência artificial é inteligente?* São Paulo: Almedina, 2023a.

SANTAELLA, Lucia. *Há como deter a invasão do ChatGPT?* São Paulo: Estação das Letras e Cores, 2023b.

SHEEHAN, Tony. Generative AI in Education: past, present, and future. *Educause Review*, 11 set., 2023. Disponível em: <https://er.educause.edu/search?q=Tony+Sheehan&sortBy=relevance&sortOrder=asc&page=1>. Acesso em: 15 fev. 2024.

TEO, Kai Xiang. OpenAI has quietly changed its core values and being 'thoughtful' and 'audacious' no longer makes the cut. *Yahoo Life*, 17 out., 2023. Disponível em: <https://ca.style.yahoo.com/news/openai-quietly-changed-core-values-042135276.html>. Acesso em: 15 fev. 2024.

UNESCO. *Recommendation on the Ethics of artificial intelligence*. Paris: UNESCO, 2022. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000381137>. Acesso em: 18 abr., 2024.

# Dilemas éticos da inteligência artificial para um ensino em metamorfose

Luis Eduardo Viegas<sup>1</sup>

**Resumo:** Diante dos desafios existentes, é necessário buscar novos modelos de ensino superior que incorporem as novas tecnologias. Conectividade, acessibilidade, plataformização e ubiquidade serão pilares fundamentais, mas também é essencial manter como farol a promoção de instituições comprometidas com a crítica, legitimadas nos campos envolvidos, responsáveis a desafios em diferentes esferas, representativas, produtoras de conhecimento rigoroso e indutoras de desenvolvimento socioeconômico. Peças fundamentais dessa transformação são as questões éticas e modelos de governança atentos em potencializar seus benefícios sem deixar de olhar para os riscos associados.

**Palavras-chave:** Inteligência Artificial Generativa; tecnologia; educação; ensino superior; dilemas éticos; transformação

---

<sup>1</sup> Doutorando em Tecnologias da Inteligência e Design Digital (TIDD) pela PUC-SP, Mestre em Administração pelo IBMEC e Graduado em Informática pela UERJ. Pesquisador em IA Responsável pela Cátedra Oscar Sala do Instituto de Estudos Avançados da USP, em Tecnologias Emergentes pelo Grupo de Pesquisa Transformação Digital e Sociedade da PUC-SP, e pelo Observatório do Futuro do Trabalho. Orcid: <https://orcid.org/0009-0003-5334-021X>.

## **Ethical dilemmas of Artificial Intelligence for an education in metamorphosis**

**Abstract:** Given the existing challenges, it is necessary to seek new models of higher education that incorporate new technologies. Connectivity, accessibility, platformization and ubiquity will be fundamental pillars, but it is also essential to maintain as a lighthouse the promotion of institutions committed to criticism, legitimized in the involved fields, responsive to challenges in different spheres, representative, producers of rigorous knowledge and inductors of socioeconomic development. Fundamental pieces of this transformation are ethical issues and governance models willing to maximize its benefits while paying attention to the associated risks.

**Keywords:** Generative Artificial Intelligence; technology; higher education; ethical dilemmas; transformation

Há um razoável consenso entre acadêmicos e especialistas sobre a necessidade de promover revisão do ensino superior como consequência das transformações ocorridas nas últimas décadas e, sobretudo, do impacto sistêmico provocado pela Revolução Digital, fundamentada na hiper conectividade da sociedade, em grande capacidade computacional, algoritmos sofisticados e, hoje, na dataficação de todos os setores da sociedade e economia.

Schwab (2016, p. 13) chama a atenção para a velocidade das transformações, a combinação de várias tecnologias e as mudanças de paradigmas sem precedentes como características destes novos tempos. Kissinger *et al.* (2021, p. 3) alertam que “a promessa de transformações, que marcam época – na sociedade, economia, política e política externa, traz o presságio de efeitos além do foco tradicional de um autor ou campo de estudo”. Isso vem intensificando efeitos que já se manifestavam nas últimas décadas como um estreitamento entre fronteiras culturais e geográficas, o questionamento sobre instituições e sistemas sociais e políticos, a concentração de renda e desigualdade social crescentes e o deslocamento da economia para um modelo predominantemente caracterizado como, segundo Dowbor (2020, p. 23), informacional e com outro modo de produção.

A disseminação das Inteligências Artificiais Generativas (IAGs) há cerca de um ano ilustra um universo de possibilidades que poderão transformar a forma como vivemos. Segundo Andre (2024), o volume de dados a serem criados e consumidos no mundo em 2024 será o dobro do que foi em 2021. Isso indica que ainda há muito por acontecer e, como destacado por Lee (2019, p. 3), estamos com muitas perguntas sem respostas, tentando prever o futuro sob uma mescla de encanto infantil com preocupações adultas em relação ao entendimento sobre nossos propósitos.

Na educação superior não é diferente. E na ausência de um oráculo ao qual se possa recorrer sobre como será o ensino superior no futuro e o que deve ser considerado para potencializar as oportunidades e conter os desafios inerentes à evolução tecnológica, são considerados como balizadores neste texto duas fontes que tratam do tema de maneiras distintas, embora ambas legítimas.

Almeida *et al.* (2023, p. 130) traçam um valioso panorama sobre fundamentos do modelo educacional e afirmam que a era da cultura educacional baseada no livro e na transmissão hierárquica do conhecimento vem perdendo sua hegemonia ao longo de mais de um século e que,

com a expansão da cultura digital, essa perda fica cada vez mais nítida. Os autores afirmam ainda que os pilares da educação superior serão a conectividade, a acessibilidade, a plataformização online e a ubiquidade, e apontam para as seguintes premissas sobre o conceito de Universidade para sua análise:

- (i) Instituição inerentemente comprometida com a crítica, incluindo a autocrítica.
- (ii) Responde a funções não apenas científicas, mas também econômicas, sociais, culturais, artísticas e éticas, devendo legitimar-se em todos esses campos.
- (iii) Deve responder, simultaneamente, a desafios em esferas locais, nacionais, internacionais e universais.
- (iv) Pertence à sociedade, à humanidade, não devendo ser submetida a interesses não públicos.
- (v) Abriga e produz conhecimento rigoroso e trabalha com o desenvolvimento da inteligência, voltada para o progresso econômico, político e social.

Já Cawood e Vasques (2022, p. 4) tratam do tema sob a ótica da viabilidade do modelo de negócios sob a premissa de que as pessoas querem aprender a partir de qualquer lugar. Os autores apoiam-se no modelo desenvolvido pela firma global de consultoria EY, que explora como a convergência de tecnologias, as mudanças demográficas e os novos modelos de negócio podem mudar a estrutura do setor educacional. Os autores segmentam as instituições de ensino entre tradicionalistas, que afirmam que seus modelos de negócios são seguros mediante preços seguidamente crescentes e perfis de procura, e os revolucionários com argumentos de que o modelo econômico atual está sob pressão existencial como decorrência de queda nas taxas de natalidade, da pressão sobre a acessibilidade, dos custos e benefícios da digitalização e dos novos concorrentes emergentes. Embora mencionem a qualidade da educação como fator relevante, o que concretamente representa qualidade da educação como parte de um novo modelo não é aprofundado.

Embora muito diferentes, as duas abordagens incorporam o uso de tecnologias emergentes como ingrediente fundamental coexistindo com outros componentes, cada qual à sua maneira, para uma nova visão de educação. E esta nova visão requer rediscutir o propósito do ensino superior e, conseqüentemente, as possibilidades que podem ser incorporadas para promover um modelo efetivo e, principalmente, compatível e coerente com o propósito das instituições.

Almeida *et al.* (2023, p. 148) chamam a atenção para o fato de que as modalidades mutantes “dos meios não podem se sobrepor aos imperativos dos fins da educação e da vida da humanidade”. Por isso, “os grandes temas humanos, que se debatiam nos anos finais do século XI, continuarão a merecer os debates e razão da vida universitária, sem se limitar às suas finalidades como provedoras de especialistas para atender a um mercado de trabalho alheio aos problemas gerais da sociedade”. Os autores enfatizam que o propósito da universidade é ser uma instituição universal, autônoma e promotora de pensamento livre e crítico. Defendem, portanto, que a metamorfose da universidade exige referências radicais, compatíveis com a intensidade dos desafios existentes e dos que ainda virão pela frente. Como referências radicais, citam a universalidade do conhecimento, da instituição e das profissões de docência e pesquisa, além do “princípio da inter e transdisciplinaridade com a devida abertura às incertezas e errância inerentes ao processo de produção do conhecimento e das artes, o que implica uma boa margem de indisciplinaridade”, o que é um indiscutível farol a guiar os movimentos diante um mar de possibilidades, muitas desconhecidas, intensificadas a partir da Revolução Digital.

### **Inteligência Artificial Generativa (IAG)**

O caráter especulativo fica ainda mais nítido quando se faz um exercício concernente aos modos de incorporação das tecnologias aos modelos de educação. A IAG, que certamente foi a tecnologia mais onipresente em 2023, vem sendo usada também por docentes e discentes. Ela é nova e age de maneira diferente dos modelos de IA difundidos até então, com predominância de finalidades preditivas e de identificação de padrões. Os modelos de IAG, por sua vez, chegaram causando grande ruído pela capacidade de gerar conteúdos, principalmente textos e imagens, com respeitável grau de coerência e efetividade.

Apesar do frenesi inicial e de bem-sucedidas experiências, surgiram muitas manifestações de resistência e de apreensão, preocupadas em como garantir um uso eticamente responsável e maduro dessa tecnologia. De um lado, desejada por uma parcela da comunidade acadêmica, de outro, pressionada por outro grupo que chegou a defender seu banimento como medida para conter os riscos associados que não tardaram a surgir. Entre os riscos levantados pela comunidade acadêmica Yu (2023, p. 3) cita: desonestidade acadêmica; comportamentos de trapaça; dificuldade para docentes determinarem se os discentes usaram modelos de IAG em

suas atividades; perda da capacidade de pensar criticamente, de explorar, verificar e sumarizar ativamente; prejuízo no processo de aprendizagem e desenvolvimento dos alunos; integridade acadêmica; e qualidade da educação.

Cerca de cinco meses após a disponibilização do ChatGPT ao público em geral, muitos intelectuais e pessoas influentes no mundo da tecnologia chegaram a assinar uma carta aberta (Pause Giant, 2023) em favor de uma pausa imediata de, pelo menos, seis meses nos treinamentos de sistemas de IA mais potentes que o GPT, argumentando que sistemas de IA avançada deveriam ser desenvolvidos apenas quando estivermos confiantes que seus efeitos serão positivos e que seus riscos serão gerenciáveis. A carta alerta sobre a existência de uma corrida descontrolada para desenvolver e implementar mentes digitais cada vez mais potentes que ninguém – nem mesmo seus criadores – consegue compreender, prever ou controlar de maneira confiável. Além disso, defende que a pesquisa e o desenvolvimento de IA devem ser redirecionados para tornar os sistemas atuais mais precisos, seguros, interpretáveis, transparentes, resilientes, alinhados, confiáveis e leais. Mesmo alguns dos signatários, como Yuval Noah Harari, admitiram que não esperavam efetiva suspensão, mas um alerta para a sociedade em geral mediante a repercussão do manifesto.

Passado um ano da popularização das IAGs, fica evidente que o mais óbvio é aceitar a coexistência entre a IAG e modelos tradicionais nos diferentes segmentos da sociedade, sem menosprezar os riscos efetivos. Santaella (2023, p. 110) conclui que os modelos de IAG funcionam como actantes, no sentido da teoria ator-rede de Latour (1998), portanto, a prioridade é compreender o que eles nos fazem fazer, o que nos fazem sentir e os modos como nos fazem agir. A autora (*ibid.*, p. 112) destaca que “é mister atender aos sinais de alerta, de olhos bem abertos e compreensão bem azeitada, pois os dados estão lançados”.

Na educação, assim como em outros setores da sociedade, inelutavelmente as IAGs são utilizadas, principalmente dada a facilitação que apresentam para interações do tipo comando (prompt) e resposta tanto em texto quanto em imagem, dependendo do sistema que é acionado. Técnicas para gerar comandos foram desenvolvidas como mecanismos para melhorar a qualidade das respostas das IAGs, tendo até mesmo surgido a pseudo-profissão de “Analista de Prompts”.

Com a sofisticação dos modelos e o desenvolvimento de soluções específicas para cada setor da sociedade, certamente as formas de uso serão plurais. Mas, para aquilo que interessa a este manual, levanta-se a

pergunta: como as IAGs serão incorporadas aos modelos educacionais? Como desenvolver um modelo mais abrangente, com capacidade de gerar conteúdos individualizados para atender às necessidades dos diferentes discentes, sem perder a fidelidade ao propósito da instituição, sob a tutoria e curadoria de profissionais capacitados? O excesso de flexibilização poderá prejudicar o desenvolvimento de certas capacidades fundamentais ligadas às carreiras?

A especulação sobre o modelo de educação do futuro, com uma boa capacidade de abstração, é muito importante para refletir sobre possibilidades, benefícios, possíveis desafios e, principalmente, como pavimentar uma estrada que permita o desenvolvimento de alternativas de maneira responsável e coerente. Isso é fundamental já que a IA gera a possibilidade de potencializar riscos e impactos que já existiam em menor medida no modelo de educação tradicional. Gera também a chance de trafegar de maneira sutil na fronteira imaginária em que se misturam, em uma zona cinzenta, o responsável e o irresponsável, o ético e o antiético, o uso produtivo e o uso perverso.

### **As exigências da ética**

WEX, o dicionário e enciclopédia jurídica do Instituto de Informação Jurídica da Faculdade de Direito Cornell, traça a origem da palavra *ética*, que deriva da palavra grega *ethos* (caráter). “No contexto jurídico, a ética define como os indivíduos escolhem interagir uns com os outros. Na filosofia, a ética define o que é bom para o indivíduo e para a sociedade e estabelece a natureza dos deveres que as pessoas devem a si mesmas e umas às outras” (*ibid.*). A publicação esclarece que, apesar de o direito muitas vezes incorporar princípios éticos, o direito e a ética estão longe de ser coextensivos: “Muitos atos que seriam amplamente condenados como antiéticos não são proibidos por lei – mentir ou trair a confiança de um amigo, por exemplo. E o contrário também é verdade. Na maior parte daquilo que a lei faz é não se limitar a codificar normas éticas” (*ibid.*). Muitas vezes, as questões éticas são tratadas como discussão teórica ou abstrata, gerando barreiras para a sua compreensão prática e consideração na esfera de atuação de indivíduos e organizações. Vencer esse obstáculo é necessário para que a ética seja incorporada como ingrediente elementar em tudo o que envolve as tecnologias emergentes.

Santaella (2021, p. 112) alerta que “quando se fala em ética, de saída, é preciso demarcar o campo. Uma vez que se trata de uma questão crucial que afeta nosso estar no mundo, nossas relações com o outro e nossos

modos de agir, existe uma ética filosófica, reflexiva e uma ética prática”. A primeira existe para iluminar a prática, quer dizer, “fornecer princípios que a prática deveria ter como alvo colocar em ação. É providencial aqui a palavra ‘ação’ pois a ética lida com o agir humano.” A autora afirma também que o poder está subjacente aos debates sobre os modos pelos quais as corporações e países criam políticas sobre desenvolvimentos e usos de tecnologias. Está igualmente subjacente às conversações sobre a democratização, equidade e responsabilidade na IA. Por isso mesmo, o debate sobre ética é central nessas questões tanto quanto é central nos diálogos entre as organizações governamentais, não governamentais, as empresas gigantes de tecnologia com seus impactos da IA na vida social e hoje, sem escusas opacas, é questão a que a educação está longe de poder escapar.

Coeckelbergh (2020, p. 7) afirma que a ética da IA trata da mudança tecnológica e do seu impacto na vida dos indivíduos, mas também das transformações na sociedade e na economia. Blackman (2022, p. 2), por sua vez, pondera que, diferentemente do que muitos cientistas e engenheiros pensam, a ética não é algo ‘mole’, pois há muito o que se fazer com ela, começando por articular objetivos, elaborar estratégias e implantar táticas para realizar essas estratégias. O autor argumenta que se deve buscar IA para aquilo que não é ruim, em vez de ‘apenas’ buscá-la para o bem, contanto que se evitem armadilhas éticas na busca de objetivos, sejam esses objetivos eticamente admiráveis ou eticamente neutros. O autor acrescenta que, em geral, as pessoas sabem como não fazer coisas ruins. Entretanto, no caso da IA, riscos éticos não são apenas consequências de mau comportamento, mas também o resultado de não se vislumbrarem possíveis consequências ou não se monitorar a IA de maneira efetiva.

Apesar dos riscos éticos da IA não serem novos (como, por exemplo, discriminação ou invasão de privacidade), ela adiciona novas vias de incorrer nesses riscos, de potencializar ou mesmo de gerá-los de maneira sutil, às vezes imperceptível. Seja pela intensidade, instantaneidade, abrangência e dimensão com que os eventos ocorrem no meio digital. E isso demanda novas maneiras de evitar ou mitigar esses riscos. Coeckelbergh (2020) argumenta que a maioria dos problemas éticos, tais como vieses, usualmente não são intencionais por parte de desenvolvedores e usuários. Uma das principais causas encontra-se na utilização de bases de dados que já carregam vieses, ou ainda qualquer outra distorção capaz de gerar problemas éticos. Ainda outra razão frequente é a adoção de premissas ou de aspectos culturais específicos que possam, ainda que de maneira involuntária, gerar problemas éticos nos resultados dos mo-

delos de IA. Não menos importante é o grau de agenciamento ou autonomia que pode ser atribuído a modelos de IA.

Quando se fala em ética na IA, três elementos aparecem recorrentemente para a maioria de autores e especialistas no tema: vieses, explicabilidade e privacidade. Embora não sejam os únicos riscos, eles são naturalmente os desafios com os quais se lida atualmente com maior frequência. Há certamente outros riscos éticos que podem advir da falta de confiabilidade, transparência, equidade, justiça, privacidade e inclusão que merecem atenção durante o desenho, desenvolvimento, implementação e acompanhamento dos modelos de IA. Há um claro desafio em mitigar esses riscos dada a complexidade de se prever as consequências que a utilização da tecnologia pode gerar. Santaella afirma que “a rigor, todos os conceitos relativos à ética são difíceis de definir e precisar”, pois não permitem uma compreensão *prima facie*. “De fato, são conceitos que ocupam centenas de páginas de discursos filosóficos e sociológicos que tentam explicitá-los, o que sinaliza que, nesse terreno, todo cuidado é pouco” (Santaella, 2021, p. 123). A autora também afirma que, em relação ao presente e, sobretudo, ao futuro, “além de vários outros quesitos éticos nas mãos dos desenvolvedores e empreendedores, as consequências dependerão dos usos, guiados ou não pela ética, que os humanos empreenderão com os sistemas” (Santaella, 2023, p. 110).

### **Por uma IAG responsável e efetiva no ensino superior**

Pimentel e Carvalho (2023) sugerem que o uso do ChatGPT requer algumas competências necessárias para que perguntas coerentes lhe sejam feitas. É preciso duvidar do conteúdo das respostas, validar certas informações fornecidas pelo modelo através de fontes confiáveis e fundamentar as informações geradas. Adiciona-se um ingrediente importante: os estudantes deverão desenvolver consciência e maturidade para melhor usufruir e se beneficiar das facilidades e potencialidades oferecidas por essa tecnologia.

Negar a tecnologia não é uma opção legítima. Promover a coexistência harmônica, sim. Para isso, o propósito do ensino superior precisa estar muito claro e haver sintonia entre os diferentes setores e as diferentes instituições, ainda que possam existir nuances distintas no propósito de instituições, de acordo com a região ou aspectos culturais. À luz do alvo da universidade proposto por Almeida *et al.* (2023), é imperativo conside-

rar a necessidade das referências às questões éticas para abrir o caminho daqueles que moldarão a universidade do futuro.

É necessário, portanto, que haja consonância com a representatividade dos diferentes segmentos da população e, sobretudo, predomine o desenvolvimento da coletividade e não de indivíduos ou grupos (sociais ou econômicos) em particular. Pode-se dizer que este deveria ser o marco zero para permitir que a universidade seja efetivamente uma instituição universal, produtora e disseminadora de pensamento livre e crítico. Também devem ser contempladas na pauta outras questões importantes como justiça, transparência, explicabilidade, respeito, responsabilização, confiabilidade, privacidade e segurança. E muito além de tratá-las individualmente, é crítico que sejam consideradas como atributos de cada componente do modelo educacional, formando uma estrutura sólida, mas ao mesmo tempo resiliente, capaz de perseguir seu propósito com atenção e abertura à dinâmica dos tempos, às oportunidades e aos riscos inerentes.

Além de serem fundamentais para se manter o propósito, as questões éticas também são a chave para oferecer um o ensino superior atraente para as novas gerações e, com isso, estimular o conhecimento, a sensibilidade diante de um mundo cada vez mais complexo e integrado, além de atender à inquietação e à autonomia das pessoas. Isso requer uma estrutura forte, com princípios educacionais e éticos sólidos e uma governança atenta para estimular e acompanhar a adequação do modelo a esses princípios. Não apenas mediante execução de padrões e critérios estabelecidos, mas também vigilante para detectar situações não previstas e reagir rapidamente, tomando as ações necessárias para corrigir rumos, pois não se pode esperar diante de um mundo em metamorfose.

## Referências

ALMEIDA, Fernando; CASALI, Alipio; DOWBOR, Ladislav *et al.*. Os desafios atuais da universidade: o mundo contemporâneo, as redes de conhecimento e o compromisso da universidade. *TECCOGS – Revista Digital de Tecnologias Cognitivas*, n. 26, jul./dez. 2022, p.129-149.

ANDRÉ, Louie. 53 important statistics about how much data is created every day. *Finances Online Reviews for Business*, atualizado 19 mar. 2024. Disponível em: [https:// financesonline.com/how-much-data-is-created-every-day/](https://financesonline.com/how-much-data-is-created-every-day/). Acesso em: 22 dez. 2023.

BLACKMAN, Reid. *Ethical machines: your concise guide to totally unbiased, transparent, and respectful AI*. Boston, Massachusetts: Harvard Business Review Press, 2022.

CAWOOD, Richard; VASQUES, Caroline. *Are universities of the past still the future?* EY, mar, 2022. Disponível em: [https://www.ey.com/en\\_gl/education/are-universities-of-the-past-still-the-future](https://www.ey.com/en_gl/education/are-universities-of-the-past-still-the-future). Acesso em: 22 dez., 2023.

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

DOWBOR, Ladislau. *O capitalismo se desloca: Novas arquiteturas sociais*. São Paulo: Edições SESC São Paulo, 2020.

KISSINGER, Henry; SCHMIDT, Eric; HUTTENLOCHER, Daniel. *The age of AI and our human future*. London: John Murray, 2021.

LATOUR, Bruno. *On recalling ANT*. In: Actor-Network Theory and After. J. Law; J. Hassard (eds.). Oxford: Blackwell, p. 15-26, 1998.

LEE, Kai-Fu. *Inteligência artificial: como os robôs estão mudando o mundo, a forma como amamos, nos relacionamos, trabalhamos e vivemos*. Tradução de Marcelo Barbão. Rio de Janeiro: Globo, 2019.

PAUSE GIANT AI Experiments: An open letter. *Future of Life Institute*. 2023. Disponível em: <https://futureoflife.org/open-letter/pause-giant-ai-experiments>. Acesso em: 15 fev. 2024.

PIMENTEL, Mariano; CARVALHO, Felipe. Implicações e teorizações dos usos das IA generativas na Educação. *Revista Cult*, São Paulo, Ed 297, p. 20-22, Setembro, 2023.

SANTAELLA, Lucia. Desafios e dilemas da ética na inteligência artificial. In: GUERRA FILHO, Willis S. *et al.* (org.). *Direito e Inteligência Artificial: fundamentos*, v. 1 – Inteligência Artificial, ética e direito. Rio de Janeiro: Lumens Juris, 2021. p. 109-136.

SANTAELLA, Lucia. *Há como deter a invasão do ChatGPT?* São Paulo, Estação das Letras e Cores, 2023.

SCHWAB, Klaus. *A Quarta Revolução Industrial*. Tradução de Daniel Moreira Miranda. São Paulo: Edipro, 2016.

WEX. *Legal Information Institute*, Cornell Law School, Ithaca. Disponível em: <https://www.law.cornell.edu/wex>. Acesso em: 27 dez. 2023.

YU, Hao. *Reflection on whether chat GPT should be banned by academia from the perspective of education and teaching*. *Front. Psychol.* 14:1181712, 2023.

# Confusões e dilemas da antropomorfização das inteligências artificiais

Anderson Röhe<sup>1</sup>

Lucia Santaella<sup>2</sup>

**Resumo:** Uma das principais razões de confusões, que rondam a compreensão daquilo que a Inteligência Artificial Generativa (IAGs) é ou não capaz de realizar, consiste em tomá-las à imagem e semelhança dos humanos, como se fossem, inclusive capazes de sentimentos ou como se elas próprias fossem racistas. Diante disso, este artigo visa debater quais os limites das IAGs e qual o papel que os humanos desempenham no uso que fazem delas. O resultado esperado é desmistificar a eficácia das IAs para a detecção de emoções ou para comportamentos éticos, sobretudo em relação aos cuidados a serem tomados em ambientes vulneráveis como a sala de aula.

**Palavras-chave:** Inteligência Artificial Generativa; emoções; racismo; uso humano

---

<sup>1</sup> Advogado na Comissão Especial de Privacidade, Proteção de Dados e Inteligência Artificial da OAB-SP. Fellow no Think Tank ABES (GT de Inteligência Artificial). Pós-graduado em Direito Digital pela Universidade do Estado do Rio de Janeiro (UERJ) e Instituto de Tecnologia e Sociedade de Rio (ITS Rio). Mestre em Políticas Internacionais pela PUC-Rio. Doutorando em Tecnologias da Inteligência e Design Digital da PUC-SP. Orcid: <https://orcid.org/0000-0002-3104-6365>.

<sup>2</sup> É pesquisadora IA do CNPq, professora titular na pós-graduação em Comunicação e Semiótica e em Tecnologias da Inteligência e Design Digital (PUC-SP). Doutora em Teoria Literária pela PUC-SP e Livre-docente em Ciências da Comunicação pela USP. Publicou 56 livros e organizou 33, além da publicação de quase 500 artigos no Brasil e no exterior. Recebeu os prêmios Jabuti (2002, 2009, 2011, 2014), o prêmio Sergio Motta (2005) e o prêmio Luiz Beltrão (2010). Orcid: <https://orcid.org/0000-0002-0681-6073>.

## **Confusions and dilemmas with the anthropomorphization of artificial intelligences**

**Abstract:** One of the main reasons for confusion, which surrounds the understanding of what Generative Artificial Intelligence (IAGs) is or is not capable of achieving, consists of taking them in the image and likeness of humans, as if they were capable of feelings or as if they themselves were racist. This article aims to debate the limits of IAGs and the role that humans play in their use of them. The expected result is to demystify the effectiveness of AIs for detecting emotions or ethical behaviors, especially in relation to the precautions to be taken in vulnerable environments such as the classroom.

**Keywords:** Generative Artificial Intelligence; emotions; racism; human use

## Introdução

A dificuldade, que o ser humano encontra em reconhecer que há outras formas de inteligência não humanas, constitui-se em uma das possíveis explicações para as tendências de antropomorfizar as Inteligências Artificiais (IAs) até o ponto de emprestar-lhes ou exigir-lhes uma consciência moral, entre outros pressupostos, como supor que os sistemas de IA sentem, mentem ou sejam racistas. Essas tendências acentuaram-se agudamente, depois do advento da IA generativa (IAG) na capacidade que ela revela de responder a demandas humanas, por verbo, imagem ou por ambos, com prontidão e até gentileza.

Há muita contradição aí: não sendo capaz de reconhecer uma forma de inteligência distinta da sua, mas não podendo deixar de perceber que os resultados das operações da IA apresentam sinais de inteligência, o ser humano se livra do problema seguindo por um entre dois caminhos: ou declara que a IAG é burra, um mero papagaio estocástico, que não entende nada do que escreve e produz ou, então, esperam-se dela reações cognitivas tipicamente humanas. Mesmo que seja verdade que a IAG não entende o significado daquilo que está escrito ou das imagens que produz, pois não tem consciência nem de si, nem consciência geral, não é isso que importa, pois ela responde e age como se tivesse. Eis a questão.

Essas contradições têm atravancado muito a compreensão da natureza e das consequências pragmáticas, ou seja, dos efeitos sensíveis que a IA tem provocado e deverá provocar, prejudicando o encontro de ações mais éticas e eficazes de lidar com ela. Por isso, no que se segue colocaremos em discussão os três fatores-chave das distorções que a antropomorfização da IA tem provocado.

### As IAGs sentem?

Inicialmente definia-se IA da perspectiva estritamente racional e só mais recentemente a pesquisa expandiu-se para o campo dos sentimentos e das emoções. Até então a questão do sentir não fazia parte do repertório, pois esbarra no “duro problema da consciência”, no sentido de ser muito complexo explicar o que é consciência (Chalmers, 2018, p. 6).

Para Tenenbaum (2018, p. 482-483), o termo “consciência” possui significados diferentes para campos do conhecimento diferentes, não havendo unanimidade, por exemplo, entre cientistas cognitivos, filósofos e neurocientistas. Tudo depende de como a conceito de consciência é to-

mado, segundo Bostron (2018, p. 11), não havendo ainda clareza do que é realmente necessário e quais as condições para o seu desenvolvimento. Outros vão mais longe, como Marcus (2018, p. 127), ao afirmar que consciência nem sequer é um pré-requisito para o desenvolvimento de uma IA. Já Koller (2018, p. 395) dispensa consciência até mesmo para se alcançar uma IA Geral ou superinteligência (um estado hipotético em que a IA suplantaria a humana). Logo, se entre pessoas já é difícil encontrar explicações para definir consciência, quanto mais em não humanos (Santalla, 2022, p. 56-66).

De todo modo, é um campo de pesquisa que se acelera a cada dia. Se nos anos de 1990 os computadores começavam a adquirir a capacidade de expressar e reconhecer afetos para, em seguida, vir a adquirir a capacidade de ter emoções (Picard, 1995), já na década seguinte, Rosalind Picard, em *Computação Afetiva* (2000), afirmava que as emoções desempenham papel mais relevante do que se imaginava; não só na tomada de decisões, mas também na percepção e no aprendizado, a ponto de influenciar os próprios mecanismos do pensamento racional. Portanto, deveria ser dado também aos computadores a capacidade de reconhecer, compreender e até mesmo de ter e expressar emoções com o objetivo de interagirem melhor conosco e aprimorarem a própria inteligência humana (Picard, 1995, p. 1).

A computação afetiva despontava, então, como nova área de pesquisa com resultados promissores, sobretudo no reconhecimento das expressões faciais e síntese da inflexão da voz (*ibid.*, p. 24), uma vez que as máquinas são melhores que os humanos para reconhecer padrões e micro sutilezas que não são tão rapidamente perceptíveis por humanos, assim como são hábeis em reconhecer quando essas inflexões se relacionam com sentimentos, tais como estresse, alegria ou raiva (Somers, 2019). As máquinas são tão eficazes nessas tarefas que isto seria apenas o início da descoberta de um estado afetivo oculto, não sendo nem mesmo necessário que esses estados sejam universais em sua expressão para que um computador os possa reconhecer.

Em *As Microexpressões Faciais: uma investigação semiótica do poder de comunicação da face humana*, Alessandra S. Cejkinski (2022, p. 1) percebe o reconhecimento das expressões faciais como etapa do próprio processo de reconhecimento das emoções. A autora o faz mostrando a divergência entre o cientista Charles Darwin, para quem as emoções são inatas e universais, e a antropóloga Margaret Mead para a qual as emoções podem sofrer interferências da cultura, do ambiente e até mesmo dos processos de aprendizagem.

A “IA emocional” ou computação afetiva foi, então, dividida em três grandes áreas de estudo por Eva Hudlicka (2008), especialista em computação afetiva aplicada aos *games*: (a) o reconhecimento de emoções; (b) modelos computacionais de emoções e (c) expressão de emoções em agentes artificiais e robôs. A primeira, de reconhecimento de emoções, é a que busca traduzir emoções humanas através de uma interação mais natural entre humanos e máquinas para depois transformá-las em dados computacionais (ver Trevisan; Braga, 2022). Quanto ao item (b), relativo aos modelos computacionais de emoções, eles são introjetados para que o resultado soe como expressão de emoções, ou seja, o item (c).

### **A IAG mente?**

As IAGs têm sido acusadas de apresentarem condutas de tipos similares a mentir. Diz-se que elas iludem, desinformam e até mesmo trapaceiam para conseguir completar uma tarefa, pois não há nada mais próximo do comportamento humano do que mentir (Martins, 2023). São atribuições que geram, então, um temor ético e de alto risco (OpenAI, 2023, p. 94).

Conforme foi amplamente discutido por Santaella (2021, p. 35-54), mentir apresenta uma semântica muito peculiar. A mentira é distinta do erro, do lapso, da hipocrisia e do cinismo. Enquanto o erro e o lapso não são deliberados, na hipocrisia e no cinismo há sempre uma certa dose de deliberação, mesmo que vaga. Mas só há mentira quando houver uma intencionalidade para mentir, cujo alvo é enganar, ou seja, fazer crer em algo que trai a realidade dos fatos.

Aqui é importante diferenciar o propósito da intencionalidade. Propósito é uma ação dirigida para um fim, agenciada pelos meios que lhe são necessários para chegar à sua finalidade. Disso são capazes animais superiores ou inferiores e, certamente, a IA, alimentada que é por estatísticas sofisticadas. Intencionalidade, entretanto, refere-se estritamente à versão psicológica do propósito. Portanto, simples assim: a IAG tem propósito, mas não tem intencionalidade, pois a intencionalidade depende da existência da consciência e autoconsciência, algo que a IA está longe de ter. Por isso, na escolha dos meios para atingir seus fins, o ser humano revela-se ético ou trapaceiro em todas as suas variedades fortes ou fracas. De resto, muitas vezes, a falta de ética pode provir de uma ingenuidade movida a ignorância ou mesmo por narcisismo que oblitera suas próprias fraquezas, ou seja, por falta de autoconsciência dos limites de si e do

contexto. Esses são comportamentos complicados que a matemática de é feita a IA não lhe permite, portanto de que a IA não poderia ser capaz. Entretanto, os bancos de dados, o treinamento e a estatística dos algoritmos podem não funcionar muito bem e a IAG produzir um resultado que fica perto, mas não chega a ser algo que alcance a psicose humana, ou seja, o ChatGPT alucina. Vejamos.

O grande diferencial do ChatGPT está na democratização de seu chatbot, fazendo com que seu sistema conversacional gere uma imediata conexão simuladamente afetiva entre usuário e máquina (Lerobitcast, 2023). Mas isso não deveria enganar o usuário, pois não dota o ChatGPT de intencionalidade e consciência humana. Consequentemente, afetividade temos que buscar em outras searas estritamente humanas, mesmo que não sejamos bem-sucedidos em nossas buscas. Tanto é que nem sempre somos bem-sucedidos que já existem sistemas de IA (Replika) para relacionamentos amorosos, basta inscrever-se, pagar uma taxa e saborear essa experiência.

Mesmo que saiba fingir, digamos assim, a IA está equipada para gerar respostas com muita prontidão, ainda que incompletas, imprecisas ou inteiramente falsas, mas evidentemente sem ter a menor noção do que é a verdade, aliás um tema que tem ocupado os filósofos há séculos. Conclusão, o propósito da IA e não sua intencionalidade é o de gerar uma resposta que, além de padronizada, muitas vezes pouco especializada, pode também passar longe da correção. Isso costuma ser chamado de alucinações (Malar, 2023), uma terminologia que deve ser vista, quando se trata da IA, dentro do seu próprio contexto, pois no caso de humanos, alucinação é caso de psicose, facilmente detectável pelos especialistas.

Todo discurso dispõe de profundidade, ou seja, sua capacidade conotativa, e de extensão, seu potencial de aplicação a situações e contextos. Ora, a IAG pode até ser capaz de ensaiar, com algum êxito, linguagens poéticas e imagens metafóricas. Mas é desprovida de senso comum, da sabedoria baseada na experiência vivida, do conhecimento tácito e outras tantas faculdades cognitivas que concedem aos humanos seus encantos e desencantos. Portanto, é justamente a carência de todas as faculdades, de certa forma parentes do senso comum, e especialmente a ausência de discernimento em saber se a resposta gerada está certa ou incorreta que esses sistemas alucinam sem nenhum apreço pelas suas fontes de referência.

Em suma: a IA não está programada para dizer a verdade, e sim “gerar conteúdos coerentes para os humanos a partir da base de dados

usada no seu treinamento” (Malar, 2023, p. 2). Portanto, dizer que a IAG não é responsável pelo que apresenta não é a mesma coisa que negar que ela possa ser usada pelos humanos para mentir, enganar, tripudiar e todos os outros tipos de malidicências que hoje encontram nas redes da web um fluxo promissor para disseminar. A IAG, portanto, não dispõe da intencionalidade para mentir, mas pode ser usada para isso. Isso também não significa negar que os bancos de dados podem estar viciados, contaminando o que se seguirá. Também não significa negar que os desenvolvedores podem ser negligentes em seus cuidados com a ética. Por fim, não significa negar que os resultados que a IA passa à frente podem servir para atender interesses escusos das mais diversas espécies.

### As IAs são racistas?

Já na metade da década de 1990, Bayat Friedman e Helen Nissenbaum (*apud* Kaufman; Junquillo; Reis, 2023, p. 44), pioneiras nos estudos sobre sistemas de computação tendenciosos, alertavam para suas ameaças à sociedade e seus potenciais impactos para os direitos humanos e fundamentais. Para tanto, identificaram três categorias de viés desses sistemas: a preexistente, cujas origens estão nas “instituições, em práticas e atitudes sociais” (Kaufman; Junquillo; Reis, 2023, p. 44), mas não necessariamente na tecnologia *per se*; a técnica (que surge de restrições operacionais dos próprios sistemas); e a emergente (a partir de determinadas situações e contextos de uso da nova tecnologia).

Portanto, um dos problemas mais sérios da IA encontra-se nos vieses, considerando-se que muitos deles precedem as IAs e podem estar presentes nos dados. Ora, dados não são neutros, e, portanto, não confiáveis. Vieses já se encontram dentro da base de dados que é utilizada para o treinamento dos algoritmos, que, em algumas situações ou contextos específicos, se revelam tendenciosos, já que alimentados predominantemente por um público masculino, de pele clara (Kaufman; Junquillo; Reis, 2023, p. 44-45) sem a devida representatividade e diversidade, seja socioeconômica, racial, étnica, regional ou de gênero. Uma “fórmula” que já se mostrou problemática consiste no algoritmo que foi considerado racista por rotular um determinado grupo de fotos de pessoas negras como sendo a de gorilas. E, quando “corrigido”, o sistema passou a não identificar nem mesmo os gorilas (*ibid.*, p. 45). Assim, se os dados não forem higienizados, ao passar pelos algoritmos e chegar aos resultados, a IA os intensifica, como bem exemplifica o racismo estrutural no Brasil.

Os problemas se acentuaram com as IAGs que automatizam a produção de textos e imagens cada vez mais realistas produzindo sérias dissociações entre fantasia e realidade. Além disso, como a maioria dos dados é produzido fora do Brasil, baseados em outras culturas, tais como a dos EUA e da União Europeia, isto é, por regiões mais ricas e poderosas do que a nossa, países representantes do Sul Global (com exceção da China, pois é um caso à parte) têm poucas chances, recursos humanos e infraestrutura computacional para tentar competir com aquelas superpotências (Santos e Soares, 2023). E, na falta de um *framework*/modelo autóctone, ao tentar replicar tais modelos, acabam não só importando problemas estrangeiros, mas também reforçando (pré)conceitos e estereótipos que já existem por aqui.

## Referências

3



# Implicações da inteligência artificial na educação

Bruno Johnson Soares<sup>1</sup>, Diego Franco<sup>2</sup>

*Colaboradores*

Bruno Sabino<sup>3</sup>, Michelle Eguchi<sup>4</sup>

**Resumo:** A inteligência artificial (IA) está promovendo mudanças significativas na educação, ressignificando as relações entre alunos e professores e levantando questões sobre a relevância futura das próprias instituições de ensino. A adoção da IA apresenta um conjunto complexo de avanços, incertezas e desafios e, apesar do potencial para aprimorar a aprendizagem e oferecer modelos inteligentes de personalização, preocupações éticas e o impacto no papel dos professores permanecem questões abertas. A integração da IA na educação, portanto, exige uma abordagem equilibrada, evitando extremos de otimismo ingênuo e pessimismo apocalíptico. A pesquisa sobre o uso ético, incluindo a responsabilidade na geração de conteúdo acadêmico e a importância do pensamento crítico, é fundamental e, universidades e instituições, carecem de diretrizes objetivas sobre o uso responsável dos modelos inteligentes, enfatizando a necessidade cada vez maior de transparência de modo que as potencialidades da IA não sejam ofuscadas pelos inerentes desafios éticos.

**Palavras-chave:** inteligência artificial; educação; preocupações éticas; transparência

---

<sup>1</sup> Bruno Johnson é mestrando em Tecnologias da Inteligência e Design Digital pela PUC-SP e especialista em Design Estratégico e Inovação pela UNISINOS. Atualmente, é Head de Produtos e Design na Compass e atua como professor convidado na ESPM/SP no curso de Design. Orcid: <https://orcid.org/0009-0005-4109-9902>.

<sup>2</sup> Diego Franco é doutorando em Tecnologias da Inteligência e Design Digital pela PUC-SP, mestre em Tecnologias da Inteligência e Design Digital pela PUC-SP, especialista em Gestão Estratégica de Marketing pela FDC. Atualmente, sua pesquisa mergulha na atuação da IA Generativa, explorando a intersecção entre tecnologia, criatividade e cultura. Também é diretor de criação e designer multidisciplinar atuando em empresas e agências de publicidade no Brasil. Orcid: <https://orcid.org/0009-0005-7379-3894>.

<sup>3</sup> Bruno Sabino é pesquisador e bacharel em comunicação social pela ESPM São Paulo. Atualmente, discente do mestrado da PUC-SP em Tecnologias da Inteligência e Design Digital com pesquisa voltada para a área de IHC na criação de imagens junto a IA. Orcid: <https://orcid.org/0009-0006-4131-8455>.

<sup>4</sup> Michelle Eguchi é pesquisadora e possui graduação em Design de Produto, especialização em Design Digital, mestrado e doutorado em Tecnologias da Inteligência e Design Digital. Sua experiência abrange a docência em UX Design, com extensa pesquisa em realidades imersivas desde 2017. Atualmente, concentra-se na linha de pesquisa em IA e Gestão, participando do comitê acadêmico da OAB-SP e contribuindo para grupos de estudo em Cidades Inteligentes e Transformação Digital. Orcid: <https://orcid.org/0000-0001-7295-4119>.

## **Implications of artificial intelligence in education**

**Abstract:** Artificial intelligence (AI) is bringing about significant changes in education, redefining relationships between students and teachers, and raising questions about the future relevance of educational institutions. The adoption of AI technology presents a complex array of advances, uncertainties, and challenges. Despite the potential to enhance learning and offer intelligent personalization models, ethical concerns, and the impact on the role of teachers remain open questions. Integrating AI into education requires a balanced approach, avoiding extremes of naive optimism and apocalyptic pessimism. Research on ethical use, including responsibility in generating academic content and the importance of critical thinking, is essential. Universities and institutions lack clear guidelines on the responsible use of intelligent models, emphasizing an increasing need for transparency so that AI's potential is not overshadowed by its inherent ethical challenges.

**Keywords:** artificial intelligence; education; ethical concerns; transparency

A emergência da IA como um fenômeno dominante iniciou uma apressada rotina de transformações em diferentes domínios da experiência humana, incluindo a educação, que se viu substancialmente sensibilizada por um repentino conjunto de inesperados avanços, incertezas e desafios. Para Kamalov *et al.* (2023), a integração da IA ao ensino está alterando a maneira como alunos e professores se relacionam e encorajando um debate recheado de interrogações sobre como as instituições podem continuar relevantes. A presença dos sistemas inteligentes na educação soa tão perturbadora quanto inevitável, reivindicando uma abordagem pragmática que renuncie tanto ao catastrofismo delirante quanto ao otimismo ingênuo.

### Questões em discussão

Ainda que a tecnologia de aprendizagem de máquina acumule funcionalidades para transformar radicalmente o sistema educacional, diversas questões permanecem pendentes. O número de incidentes éticos, afirmam Kamalov *et al.* (2023), continuou a crescer em linha com o aumento da presença da tecnologia na rotina das pessoas. Várias tentativas de diretrizes éticas surgiram recentemente, no entanto, prosseguem os autores, poucas estão direcionadas para as questões específicas levantadas pela IA na educação como estão as Diretrizes de Ética da Comissão Europeia para IA confiável e o Manual IESALC da UNESCO (Recommendations, 2021) sobre os impactos dos sistemas inteligentes no ensino superior. Embora haja reconhecimento geral sobre a relevância da IA no desenvolvimento tecno científico, falta uma abordagem politicamente abrangente e uma orientação clara em torno da IA na educação.

As consequências são multifacetadas e incluem, além da aprendizagem personalizada, sistemas de tutoria inteligentes, automação de avaliações e novos modelos de interação professor-aluno. Kamalov *et al.* (2023) também sinalizam que padrões e tendências nos dados dos estudantes podem ser identificados com a ajuda da IA e ajudar educadores a reconhecer lacunas de aprendizagem. Os autores ainda afirmam que a presença dos modelos autônomos na educação pode fornecer experiências de aprendizagem customizadas com o suporte de rastreamento de conhecimento e filtragem colaborativa. A tecnologia, ainda, pode ser usada para prever necessidades específicas dos alunos, detectando valências e limitações individuais, monitorando estados de conhecimento e analisando padrões de compreensão. Porém, para Williamson (2023), existem riscos

escondidos no entusiasmo com a presença da IA no futuro da educação. Afinal, são frágeis e atualmente insustentáveis o juízo e a crença de que modelos autônomos funcionarão conforme pretendido e planejado e que quaisquer problemas, que surjam no seu desenvolvimento, implementação e implantação, serão amenizados e resolvidos por meio de ajustes técnicos.

Para Kamalov *et al.* (2023), outro aspecto preocupante a ser examinado são os riscos do descrédito no prestígio formal dos professores, pois podem deixar de ser vistos como a autoridade final no conhecimento. A capacidade dos sistemas inteligentes de oferecerem aprendizagem individualizada, avaliações eficientes e *feedback* adaptativo impõe novos dilemas e afeta a dinâmica tradicional do vínculo educativo. O emprego generalizado da IA gera preocupações com a possível diminuição da conexão humana e erosão do relacionamento em sala de aula, o que pode abalar o desenvolvimento de habilidades socioemocionais e o vínculo de comunidade em educadores e estudantes, manifestando a necessidade, portanto, de padrões de conduta, quadros regulamentares e orientações sobre o usos éticos da IA.

A evolução da educação em direção a uma condição mais autônoma, lembram Celik *et al.* (2022), não implica necessariamente que as pessoas precisarão de menos professores no futuro. Em vez de especular se a IA irá substituir professores, o autor afirma que, compreendendo as vantagens que a IA oferece à aprendizagem, é mais razoável refletir sobre como esses recursos podem auxiliar o papel do docente em sala de aula. Na esteira do avanço da participação dos modelos de *machine learning* na aprendizagem, novas tecnologias serão apresentadas, portanto, para Kamalov *et al.* (2023), o mais importante é educar as pessoas sobre o dever de estarem atentas às demarcações éticas ao usar ferramentas e aplicações. Um exemplo recente é o lançamento da iniciativa *Frontiers of Computing* na Universidade de Sul da Califórnia (USC Frontiers, 2022), cujo objetivo é incorporar alfabetização, ética e responsabilidades digitais/IA em todas as disciplinas.

A pesquisa científica tem como objetivo fornecer caminhos e alternativas a problemáticas nas mais diversas áreas. Conforme Santaella (2016), a essência fundamental da pesquisa não reside na sua classificação como básica ou aplicada, mas sim na busca pela excelência, fundamental para evitar resultados duvidosos. A qualidade da pesquisa está intrinsecamente ligada à criatividade e habilidade de formular perguntas relevantes, sendo todas as outras fases derivadas desse aspecto crucial, e, segue a autora, construir um repertório a partir de leitura e pesquisa.

Assim, uma ferramenta de IA pode conduzir eficazmente uma pesquisa? Apesar de ter acesso a uma vasta quantidade de textos, os modelos de processamento de linguagem natural não são infalíveis. Como ressalta Santaella (*ibid.*) uma parte essencial da pesquisa é, de fato, o pensamento crítico, a seleção de referências que irão compor a pesquisa e, claro, a criatividade na construção do problema de pesquisa.

## O caso do ChatGPT

Ao avaliar o ChatGPT – sistema de IA generativa (IAG) que tem estado na crista das preocupações – para fins da pesquisa científica que resulta em conhecimento a ser utilizado como fonte nos processos educacionais, é essencial considerar critérios como referências, criatividade, senso crítico e repertório (que podemos entender aqui como leituras). Para isto, Yang, Wang e Lyu (2023) conduziram uma avaliação desses aspectos em um artigo de 2023, concluindo que o ChatGPT apresentou habilidades notáveis ao lidar com perguntas baseadas em fatos e ao fazer julgamentos precisos em questões hipotéticas, dedutivas, explicativas e de inferência simples. No entanto, os autores apontam também para o fato de que sua performance foi menos consistente em perguntas mais complexas que envolviam múltiplas etapas de inferência e correlações mais elaboradas, indicando uma limitação em sua capacidade de abordar o raciocínio lógico mais sutil e complexo. Concluem os autores que o ChatGPT tem habilidades de pensamento crítico, mas apresenta desempenho inferior quando se trata de raciocínio lógico.

Ainda mais graves são os achados de Buriak *et al.* (2023) de que o ChatGPT pode inventar referências ou criar correlações falsas. Na literatura especializada, dá-se o nome de “alucinação” a esta confabulação. De fato, modelos de LLM (processamento de linguagem natural), como o chat, geram informações que podem parecer plausíveis, mas que são, na verdade, factualmente incorretas ou não relacionadas ao contexto fornecido. Essa “alucinação” ocorre quando o ChatGPT produz respostas que não estão fundamentadas em dados reais ou na compreensão adequada do mundo (Ahmad *et al.*, 2023). A constatação de que o programa gera referências a um estudo inexistente é motivo de considerável apreensão. Esse incidente suscita inquietações sobre a confiabilidade desses sistemas e instiga reflexões críticas sobre a ética de uso no âmbito de estudos e pesquisas.

Devido a estas constatações e à grande popularidade do ChatGPT, há um acalorado debate em andamento entre editores de periódicos, pesquisadores e editores sobre o papel das ferramentas de IA, como o ChatGPT,

em publicações científicas. Após a publicação de artigos que incluíram o ChatGPT como autor, revistas de renome, como *Nature*, *Science* e *JAMA*, estabeleceram uma nova política. A primeira, e mais notável, foi a de proibir nomear ferramentas de IA como autores creditados em um artigo de pesquisa, argumentando que a atribuição de autoria implica responsabilidade pelo trabalho, algo que as ferramentas de IA não podem assumir. Além disso, essas políticas destacam que os pesquisadores são fortemente aconselhados a documentar o uso dessas ferramentas nas seções de métodos ou agradecimentos e deixar claro como e qual ferramenta de IA foi utilizada (Tools, 2023; Flanagin *et al.*, 2023; Thorp, 2023).

Na mesma linha e com as mesmas preocupações, a revista *Accountability in Research* publicou, em janeiro de 2023 (Hosseini *et al.*, 2023), um editorial colocando, também, suas considerações sobre o uso ético de Large Language Models (LLGs) na pesquisa acadêmica. Com o título de “Utilizando a inteligência artificial para escrever publicações acadêmicas”, o texto dá algumas diretrizes éticas para o uso de ferramentas como o ChatGPT, chamando atenção para a necessidade de verificar o conteúdo gerado pela ferramenta de IA por um especialista no domínio, responsabilização dos coautores em caso de erros ou vieses, divulgação e indicação do uso do LLM e de textos escritos ou coescritos por esses modelos. O texto ainda destaca a proibição do uso destas ferramentas para fabricar ou falsificar dados, reforçando a importância da integridade na condução da pesquisa e, claro, a preocupação com a alucinação dos modelos.

No entanto, esse debate não está restrito apenas às editoras. Universidades renomadas – notadamente norte-americanas – foram pioneiras em divulgar também seus guias de boas práticas, aconselhamentos e ponderações do uso destes modelos no âmbito acadêmico. Em seu site oficial, a Universidade de Harvard (Initial Guidelines, 2023) publicou um guia autoral de uso de IA na sua comunidade acadêmica onde enfatiza a importância de proteger dados confidenciais, sejam de pesquisa, sejam da própria universidade, desencorajando a inserção desses dados em ferramentas terceiras de IA. A preocupação válida reside no fato de, além de grande parte dos modelos de LLM disponíveis hoje serem privados, estes sistemas de IA aprenderem a partir das interações com os usuários e das vastas quantidades de dados gerados nestas interações. Nisso, há o claro risco do sistema inadvertidamente aprender e gerar informações sensíveis ou identificáveis pessoalmente (Sebastian, 2023).

Em consonância com as preocupações éticas para a pesquisa e consequentemente para a educação, outras universidades, como Queen’s, University of New England (Guidelines for use, s.d.), Utah e Arizona State (Guidance on the use of Ai, 2023), divulgaram seus materiais que

abrangem desde a necessidade de transparência e divulgação dos autores em explicitar o uso de IA em suas pesquisas, até a responsabilidade e prestação de contas, reconhecendo o consenso de que as ferramentas de IA não são consideradas autores e, portanto, colocam a responsabilidade nos pesquisadores para garantir a precisão científica do trabalho (artificial intelligence tools, 2023); University of North Carolina, Chapel Hill (Generative AI, 2023); Queen's University (Guidelines for, 2023; Arizona State University (Artificial Intelligence, 2024).

Outros aspectos éticos surgem. Um ponto polêmico e não consensual é sobre plágios. Algumas universidades entendem que é necessário evitar o plágio que pode vir da geração da IA. Segundo esta visão, esses modelos, como são treinados em inúmeros dados abertos (e sobre quais serão usados, até o momento, não temos visibilidade), podem plagiar conteúdos. Sobre isso, as universidades signatárias desta visão defendem a importância de práticas adequadas de citação e atribuição – que, como vimos, no caso da IA, são difíceis de serem feitas. Os especialistas dessas universidades também advogam pela supervisão humana ativa no processo de revisão de conteúdos gerados por IA reconhecendo as limitações e imprecisões inerentes a essa tecnologia (*ibid.*).

Além das considerações comuns abordadas por múltiplas universidades, algumas instituições específicas, como a Universidade de Harvard, enfatizam a importância da proteção de dados confidenciais ao usar ferramentas de IA generativa, especialmente em dados sensíveis da universidade e da pesquisa. Outras, como a Universidade de Utah, proíbem o uso de IA na revisão científica por pares devido a preocupações com a confidencialidade (*ibid.*).

### **Diretrizes assumidas pelas universidades**

Universidades como a Estadual do Arizona e a de New England estabelecem diretrizes específicas para considerações éticas ao tomar decisões durante a pesquisa, alertando contra a possível dependência do pesquisador com a ferramenta de IA e enfatizando que o uso dos resultados da IA pode vir com informações e decisões baseadas em fatores adicionais e, possivelmente, externos à pesquisa. A Universidade Estadual do Arizona, por sua vez, destaca considerações semelhantes porém relacionadas ao HIPAA (*Health Insurance Portability and Accountability Act*) em contexto de dados de saúde. Em outras palavras, essas instituições ressaltam a importância de não confiar apenas na IA para a tomada de decisões de pesquisa e incentivam, caso haja o uso de IA, a complementação dessas decisões com análises críticas sempre explicitando o processo. Além dis-

so, é importante ressaltar que a Universidade Estadual do Arizona traz fortes preocupações éticas relacionadas à privacidade dos dados em pesquisas de saúde uma vez que, muitas vezes, estes dados são coletados com indivíduos sujeitos a testes (*ibid.*).

Adicionalmente, as universidades Queen's e de North Carolina estendem esta discussão para além do âmbito da pesquisa acadêmica, abordando questões como a integridade acadêmica no ensino e os impactos nos critérios de concessão de bolsas. Para estas duas instituições, as universidades concedem bolsas de pesquisas, muitas vezes, pela originalidade das ideias do pesquisador, o que, em um contexto de IA, levanta preocupações sobre a própria originalidade do trabalho, uma vez que pode ter sido feito com alguma ferramenta de IA (*ibid.*).

Contudo, nem todas as diretrizes são proibitivas ou restritivas. Um conjunto de universidades promove e incentiva de maneira ética o emprego da IA no processo de pesquisa, reconhecendo suas qualidades. Há orientações que destacam os casos nos quais a IA pode otimizar procedimentos, como em traduções e análise extensiva de dados, ou transcrições de entrevistas; além de melhorias na revisão e edição de textos, aproveitando a IA para revisão gramatical e aprimoramento da clareza e concisão textual. Nesse contexto, a Harvard University e Queen's University reconhecem a IA como um valioso auxílio na escrita acadêmica (*ibid.*).

De maneira convergente, a University of Utah e Queen's University percebem essas ferramentas como facilitadoras do processo de pesquisa, especialmente como sumarizadores de texto que simplificam a descoberta de novos documentos e a investigação de fontes relevantes que podem adensar o conteúdo e a pesquisa. Por sua vez, a Queen's University e University of Utah atentam para a geração de ideias bem como para a estruturação de conteúdo, utilizando ferramentas de IA para desenvolver e organizar ideias de forma mais coesa (*ibid.*)

Algumas tarefas, quando do momento da publicação de uma pesquisa, podem ser facilitadas com o uso de ferramentas de IA. Neste contexto, a Harvard University e a Arizona State University sugerem o uso da IA para suporte em tarefas específicas, como redação de abstracts, análise de grandes volumes de dados e formatação de documentos para periódicos específicos – tarefa que tem baixíssimo impacto na pesquisa em si mas que demanda muito tempo do pesquisador. As diretrizes da University of North Carolina e Arizona State University consistem no reconhecimento da contribuição da IA para a criatividade e inovação, pois, segundo essa visão, a AI tem a qualidade de gerar resultados diversificados com base em

entradas específicas. Por fim, a Harvard University, a University of Utah e Arizona State University identificam a IA como uma ferramenta capaz de melhorar a eficiência de pesquisadores e acadêmicos, proporcionando suporte em várias etapas repetitivas e processuais da pesquisa acadêmica (*ibid.*).

### Súmula das diretrizes

Para todos os pontos relatados acima, estes documentos fornecem algumas boas práticas e diretrizes que podem ser praticados e incorporadas a outras comunidades acadêmicas. São diretrizes para como conduzir as discussões na comunidade acadêmica. Abaixo, listamos um pequeno compilado, não exaustivo, das práticas propostas.

(i) Promoção de discussões éticas sobre o uso de ferramentas de IA, reconhecendo a natureza “mutante” da tecnologia e a necessidade de constantemente revisar-se as diretrizes para estarem adaptadas ao momento (Guidance on the use of AI, 2023; Initial guidelines, 2023).

(ii) Estabelecimento de diretrizes claras considerando aspectos como privacidade, responsabilidade, segurança, transparência, justiça, controle e gerência humana da tecnologia e responsabilidade com a integridade acadêmica (*ibid.*).

(iii) Construção de confiança entre a equipe de ensino e os alunos, incentivando a colaboração e compartilhamento de experiências. Isto pode ser feito através de reuniões entre professores para discutir ferramentas e aplicações, com foco em usos encorajadores; grupos de discussão na comunidade acadêmica dentre outras formas. O ponto principal é não banir a tecnologia, mas rever processos de desenvolvimento de aprendizagem e entender como esta tecnologia se situa no contexto acadêmico, quais são suas qualidades e defeitos e quais podem ser os usos éticos. (Generative AI usage guidance, 2024; Guidelines for ai use, s.d.).

(iv) Estabelecimento de expectativas claras na comunidade acadêmica sobre o uso destas ferramentas, identificando práticas acadêmicas aceitáveis e inadequadas (*ibid.*).

(v) Colaboração entre pesquisadores de diversas comunidades acadêmicas para encontrar maneiras éticas de usar o ChatGPT e outras ferramentas na produção acadêmica, garantindo imparcialidade e responsabilidade (*ibid.*).

(vi) Aprimoramento e discussão sobre como esta tecnologia pode impactar positiva e negativamente na produção acadêmica e produzir materiais, guias e diretrizes que possam servir de exemplo (*ibid.*).

(vii) Reconhecimento e compreensão das limitações de cada ferramenta de IA bem como trabalhar com letramento sobre a tecnologia em todos os âmbitos e contextos da universidade (*ibid.*).

(viii) Coleta de evidências sobre o uso do ChatGPT e outras ferramentas buscando experiências dos alunos, de professores e da comunidade acadêmica, visto que é uma tecnologia de uso geral e que é ainda desconhecida (*ibid.*).

## Referências

- AHMAD, Zakia; KAISER, Wahid; RAHIM, Sifatur. Hallucinations in ChatGPT: An unreliable tool for learning. *Rupkatha Journal*, Raghunathganj, v. 15, n. 4, p. 1-18, 2023.
- ARTIFICIAL INTELLIGENCE (AI) TOOLS – Guidance on the use of AI in human subjects research. *University of Tennessee – Knoxville*, Office of Research, Innovation & Economic Development. Disponível em: <<https://research.utk.edu/research-integrity/artificial-intelligence-ai-tools/>>. Acesso em: 9 jan., 2024.
- BURIAK, Jillian M. *et al.* Best practices for using AI when writing scientific manuscripts. *American Chemical Society (ACS) Nano*, Columbus, OH, v. 17, n. 5, p. 4091-4093, 2023.
- CELIK, Ismail; DINDAR, Muhterem; JÄRVELÄ, Sanna; MUUKKONEN, Hanni. The promises and challenges of artificial intelligence for teachers: A systematic review of research. *Tech Trends*, São Paulo, v. 66, p. 616-630, mar. 2022.
- FLANAGIN, Annette; BIBBINS-DOMINGO, Kirsten; BERKWITS, Michael; CHRISTIANSEN, Stacy L. Nonhuman “authors” and implications for the integrity of scientific publication and medical knowledge. *Journal of the American Medical Association (JAMA)*, Lindale, TX, v. 329, n. 8, p. 637–639, 2023.
- GENERATIVE AI USAGE GUIDANCE for the research community. *University of North Carolina*, Chapel Hill, The Provost. Disponível em: <https://provost.unc.edu/generative-ai-usage-guidance-for-the-research-community/>. Acesso em: 9 jan., 2024.
- GUIDANCE FOR the use of artificial intelligence (AI) in research. *University of New England (UNE)*, 10/5/23. Disponível em: <<https://library.une.edu/wp-content/uploads/2023/10/Guidance-for-the-Use-of-Artificial-Intelligence-AI-in-Research.pdf>>. Acesso em: 12 jan., 2024.
- GUIDANCE ON the use of AI in research. *The University of Utah*, Office of the Vice President for Research. 13jul., 2023. Disponível em: <https://attheu.utah.edu/facultystaff/vpr-statement-on-the-use-of-ai-in-research/>. Acesso em: 9 jan., 2024.
- GUIDELINES FOR AI USE in graduate research. *Queen’s University – Graduate Studies and Postdoctoral Affairs*, s.d. Disponível em: [https://www.queensu.ca/files/uploaded\\_files.pdf](https://www.queensu.ca/files/uploaded_files.pdf). Acesso em: 12 jan. 2024.
- GUIDELINES FOR USE of artificial intelligence (AI) in research. *Arizona State University*, s.d. Disponível em: <https://researchintegrity.asu.edu/export-controls-and-security/artificial-intelligence>. Acesso em: 12 jan., 2024.

HOSSEINI, Mohamed; RASMUSSEN, Lisa M.; RESNIK, David B. Using AI to write scholarly publications. *Accountability in Research*, London, 25 jan., p. 1–9, 2023.

INITIAL GUIDELINES for the use of Generative AI tools at Harvard. *Harvard University Information Technology*. Cambridge, MA, 2023. Disponível em: <https://huit.harvard.edu/ai/guidelines>. Acesso em: 9 jan., 2023.

KAMALOV, Firuz; CALONGE, David; GURRIB, Ikhlās. New era of artificial intelligence in education: Towards a sustainable multifaceted revolution. *Sustainability*, Basel, v. 15, n. 16, 2023.

RECOMMENDATIONS ON THE ETHICS of artificial intelligence. *Unesco Unesdoc Digital Library*, Document code: shs/bio/rec-aiethics/, 2021. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000380455>. Acesso em: 9 jan., 2024.

SANTAELLA, Lucia. Há como escamotear as questões do método? *Organicom*, São Paulo, v. 13, n. 25, p. 50–68, 2016. Disponível em: <https://www.revistas.usp.br/organicom/article/view/139338>. Acesso em jan. 2024.

SEBASTIAN, Glorin. Privacy and data protection in ChatGPT and other AI chatbots: Strategies for securing user information. *Social Science Research Network (SSRN) e-library*, 25 mai, 2023. Disponível: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4454761](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4454761). Acesso em: jan. 2024.

THORP, H. Holden. ChatGPT is fun, but not an author. *Science*, Washington, DC, v. 379, n. 6630, jan. 2023, p. 313.

TOOLS such as ChatGPT threaten transparent science; here are our ground rules for their use. *Nature*, London, v. 613, n. 7945, p. 612, jan. 2023.

USC FRONTIERS in computation. *University of Southern California*, 2022. Disponível: <https://computing.usc.edu>. Acesso em: jan. 2024.

WEBB, Michael, *et al.* Artificial intelligence (AI) in tertiary education: A summary of the current state of play. *Joint Information Systems Committee (JISC) Reports*, 2ª ed. Kew: National Archives, 2022. Disponível em: <https://repository.jisc.ac.uk/8783/1/ai-in-tertiary-education-report-june-2022>. Acesso em: 24 jan. 2024.

WILLIAMSON, Ben. The social life of AI in Education. *International Journal of Artificial Intelligence in Education*, Berlin, v. 34, p. 97–104, 2024.

YANG, Xiaozhe; WANG, Qinggin; LYU, Jiyue. Assessing ChatGPT's educational capabilities and application potential. *East China Normal University (ECNU) Review of Education*, Shanghai, 5 nov. 2023.

**APREND.AI:****Inteligência Artificial  
Generativa no ensino superior**Tina Fonseca<sup>1</sup>Lucila Campiglia<sup>2</sup>

**Resumo:** O capítulo propõe um framework de design ético para a transformação do cenário educacional, abordando a colaboração entre uma instituição de ensino superior (IES) e Inteligências Artificiais (IA), especialmente a Generativa (IAG), desempenhando atividades antes exclusivas dos humanos. Diante da urgência de repensar as práticas de ensino-aprendizagem, especialmente nas avaliações tradicionais, surge o Design Ético de Modelo de Mudança, denominado *Aprend.ai*. Essa abordagem visa fornecer à comunidade acadêmica um ambiente controlado para orientar a transição, revisando competências curriculares para integrar IAG às práticas de ensino-aprendizagem, com a participação ativa de alunos e professores.

**Palavras-chave:** aprendizagem assistida por IA; avaliação educacional; modelo de mudança; design ético; meta-aprendizagem

---

<sup>1</sup> Doutoranda no Programa de Tecnologias da Inteligência e Design Digital da Pontifícia Universidade Católica de São Paulo, pesquisadora, bolsista do CNPq, jornalista e editora, docente. Consultora de modelo de mudança ético para uso de IAGs em ambientes educacionais e o uso assistivo de IAs, IAGs, RV, RA e RM no ensino profissionalizante inclusivo - [ra00341800@pucsp.edu.br](mailto:ra00341800@pucsp.edu.br). Orcid: <https://orcid.org/0000-0002-4766-1784>.

<sup>2</sup> Mestranda em Comunicação e Práticas do Consumo pela ESPM, pós-graduada em Negócios da Moda pelo Senac SP, graduada em Arquitetura e Urbanismo pela Faculdade de Arquitetura e Urbanismo da Universidade de São Paulo FAU-USP. É palestrante, docente e pesquisadora em comunicação, semiótica e tecnologia. Orcid: <https://orcid.org/0009-0002-9054-8979>.

## **ILearn.AI: Generative Artificial Intelligence in higher education**

**Abstract:** The chapter proposes an ethical design framework to transform the current educational landscape, addressing collaboration between a higher education institution (HEI) and artificial intelligence (AI), particularly Generative AI (GAI), engaging in activities once exclusive to humans. Faced with the urgency to rethink teaching and learning practices, especially in traditional assessments, as highlighted by Santaella (2023), emerges the Ethical Design of Change Model, referred to as Apend.ai. This approach aims to provide the academic community with a controlled environment to guide the transition, revising curriculum competencies to integrate GAIs into teaching and learning practices, with active participation from students and professors.

**Keywords:** AI-assisted learning; educational assessment; change model; ethical design; metalearning

Com um crescimento – comprovado – do zero a um número superior a um milhão de usuários em apenas cinco dias, é fato que o GPT (Transformador Generativo Pré-Treinado), um dos sistemas mais conhecidos da Inteligência Artificial Generativa (IAG) é tido como um dos principais catalisadores da adoção em massa das tecnologias de IA até aqui. Dado o potencial desse sistema de dar respostas prontas e até certo ponto satisfatórias ao usuário em textos sintaticamente perfeitos, esta não é a primeira, tampouco a única ameaça à integridade acadêmica. Quase três décadas atrás, em 1996, o físico norte-americano Alan Sokal publicou um artigo na revista *Social Text* sobre a hermenêutica da gravidade quântica. Seu texto parodiava um *mix* de artigos em linguagem pós-modernista e acabou ironicamente capitaneando um megadebate sobre a respeitabilidade intelectual dos estudos culturais.

Pouco mais tarde, Phillip Ball (2005) publicou um artigo relatando o caso de Jeremy Stribling (formado em ciência da computação pelo MIT) que, juntamente com Daniel Aguayo e Maxwell Krohn, havia utilizado um gerador automático de artigos para reunir textos e gráficos aleatórios, com o objetivo único de testar se esse tipo de *patchword*, neologismo criado por Fonseca (2024) para se referir a retalhos de textos não autorizados, falsos ou gerados por softwares, costurados para publicação com o intuito de burlar a Academia. Esse *patchword* passaria despercebido na triagem da 9ª Multiconferência Mundial sobre Sistêmicos, e, como resultado, o conteúdo foi aceito na Conferência Cibernética e Informática (WMSCI). Van Noorden (2014) também mencionou o caso das editoras Springer e IEEE, que removeram mais de 120 artigos de seus serviços de assinatura quando perceberam que todos não passavam de *patchwords*.

A lista é extensa, mas ninguém supera Ike Antkare, que teria produzido mais de uma centena de publicações em apenas um ano, ocupando, com isso, a 2ª posição entre os “estudiosos” mais citados do mundo – passando à frente de Einstein – como revelou Labbé (2010). Casos como esses fizeram nascer na União Europeia (EU) o Grupo de Peritos de Alto Nível sobre IA de Confiança (doravante denominadas Diretrizes do HLEG), com a missão de descrever os requisitos necessários para um desenvolvimento responsável e confiável na UE.

Agora, chegou a vez das questões de autoria e integridade da informação relativas à IAG, em especial ao ChatGPT cujo potencial criativo parece se mostrar igual ou superior ao humano. De fato, não pode deixar de ser atemorizante a emergência de uma entidade artificial provida de inteligência que pretensamente aprendeu de modo rápido tudo sobre to-

das as coisas de todo o mundo e capaz de transmiti-las de forma fluida e para alguns convincente. Em meio às variadas problemáticas que têm sido levantadas sobre a IAG, este artigo pretende lançar a proposta de um design ético de modelo de mudança, próprio para apoiar gestores, professores e alunos na transição para uma nova realidade acadêmica: a da Aprendizagem Assistida por IA.

### Uma bússola ética

Em 2023, havia mais de 92 mil artigos resultantes da busca por ChatGPT só no motor do *Google Scholar*, desses, 37.200 resultantes apenas do último ano. Curiosamente, ao acrescentarmos o termo professor, obtivemos 3.420 resultados; ética (1.580); educação (1.050); ensino superior (326); integridade acadêmica (30). Contudo, não houve resultado para as buscas incluindo as palavras-chave modelo de mudança e design ético – o que denota a importância de apresentar tais conceitos no contexto do uso responsável das IAGs no âmbito da educação superior, como uma espécie de modelo que certamente pode ser readaptado para outros níveis de educação.

Um levantamento da empresa Impact Research (*ChatGPT Used by Teachers, 2023*) identificou que mais da metade (51%) dos professores respondentes (foram 1 mil ao todo) usam o ChatGPT – sendo que 40% semanalmente e 10% quase que diariamente. Não por acaso, Santaella (2023, p. 86) classifica como uma “emergência tecnológica” diminuir o gap entre “alunos de prontidão” e “professores ainda despreparados para mudanças no planejamento”. Emergência essa que não deverá ser atendida com sugestões efervescentes, mas com treinamento dos alunos para o uso adequado dessa tecnologia, melhorando a experiência de trabalho interativo com modelos de IA. Para atender essa demanda, Fonseca (2023) apresentou no IV Congresso de Inteligência Artificial da PUC-SP a proposta de desenvolvimento do *Aprend.ai* – um ambiente controlado que considera elementos dos sistemas de recomendação para ajudar a personalizar e direcionar a mudança para o modo de Aprendizagem Assistida por IA partindo de um design ético.

Embora não haja uma definição acabada para design ético (sobretudo no que tange ao desenvolvimento de design digital), a expressão quase sempre se refere a um tipo de design que resiste a padrões manipulativos, respeita a privacidade dos dados, incentiva o *co-design* e é acessível e centrado no ser humano. Há, inclusive um consenso sobre o estabelecimento de uma base moral – ou o que Abend (2014) chama de “antecedentes morais” – e que está intrinsecamente relacionado à pesquisa interdiscipli-

nar em valores no design de tecnologia e à sociologia dos negócios e da ética profissional.

Em seu artigo “Better, nicer, clearer, fairer: A critical assessment of the movement for ethical artificial intelligence and machine learning”, Greene, Hoffman e Stark (2019) fazem uma análise detalhada de declarações de valores de alto perfil que endossam o design ético para IA, e que legitimam (e deslegitimam) certas práticas, ao fornecerem respostas para os porquês de se fazer tal coisa e de tal forma – sempre objetivando algo melhor, mais agradável, mais claro e mais justo.

A aplicação da ética antecipatória ao design de tecnologias digitais emergentes de Shilton (2018), mostra como a ética e os códigos éticos designam e defendem o *status* social e a experiência, mais do que impõem virtudes morais ou sociais consistentes. É no design ético que reside o poder de erradicar padrões obscuros, vieses, *apartheid* digital etc. (Ferreira, 2020). Como todo bom design, o *Aprend.ai* deverá observar as dez heurísticas de Nielsen (1994):

- Visibilidade do status do sistema
- Correspondência entre o sistema e o mundo real
- Controle e liberdade do usuário
- Consistência e padrões
- Prevenção de erros
- Reconhecimento em vez de lembrança
- Flexibilidade e eficiência de uso
- Design estético e minimalista
- Ajuda aos usuários para reconhecer, diagnosticar e se recuperar de erros
- Ajuda à documentação

Wickramasinghe (2020) reforça que a interação humano-computador está concentrada na experiência do usuário (UX) e considera essencial estabelecer diretrizes que garantam a conformidade de sistemas interativos com documento de requisitos, especificações, diretrizes ou características relativas à adequação à finalidade dos materiais, produtos, processos e serviços, com destaque para a adoção de um desenho universal que considere deficiência, idade e cultura.

### **A formação de professores na era da aprendizagem assistida por IA**

Para Roberto Araya (2018), do Centro de Investigação Avançada em Educação da Universidade do Chile, encontrar formas de melhorar o ensino configura o maior desafio da educação, especialmente pelo fato de que medir a qualidade do ensino é difícil e caro (em 2016 um estudo do

tipo custaria 11 milhões de dólares) – o que fez surgir uma série de alternativas para alcançar esse objetivo: a formação, programas de mentoria e sistemas de apoio para melhorar o desempenho dos professores, são algumas delas. Suas pesquisas revelaram o potencial transformador do *ConectaIdeas* – uma plataforma baseada na Web, que se baseia em metodologias ativas de ensino, mas que se comporta como uma espécie de jogo, promovendo a constante interação entre professores e alunos.

Na lógica de Ferry (2015), estamos realizando processos que, ao que tudo indica, poderão ser logo substituídos por uma IA, exceto se nós a integrarmos como um poderoso instrumental e não como o núcleo central da *Matrix*. Nosso modelo de mudança recoloca o professor no lugar de mestre, devolve seu protagonismo como mediador/condutor crítico da relação entre inteligências humanas e artificiais e com elas o qualifica para o próximo estágio na vida acadêmica interdisciplinar: mentor proficiente, curador de conteúdos, analista de *prompts*, gestor de tecnologias instrucionais, com capacidade técnica para capitanear uma produção acadêmica de mais alto nível e aplicabilidade. Mas, como a aprendizagem assistida por IA pode ser implementada com base na teoria da aprendizagem ativa? Seguem algumas indicações.

(i) Envolvimento ativo dos alunos: A Aprendizagem Ativa destaca a importância do envolvimento ativo dos alunos. Com o uso de IA, os sistemas podem criar ambientes de aprendizagem personalizados, adaptando os métodos de ensino com base no desempenho, estilos de aprendizagem e interesses individuais dos alunos.

(ii) Feedback contínuo e personalizado: Torna possível oferecer feedback contínuo e personalizado, identificando áreas específicas em que um aluno pode precisar de mais atenção ou prática. Isso está alinhado com a ideia de Life Long Learning.

(iii) Aprendizado baseado em problemas: Essa abordagem incentiva os alunos a resolverem problemas do mundo real. Aqui diferentes IAs combinadas com realidade virtual e/ou aumentada podem simular cenários práticos, criar desafios significativos e oferecer suporte durante todo o processo.

(iv) Colaboração e interação: Uma IA pode facilitar a colaboração e interação entre os alunos, promovendo o aprendizado social. Um ambiente favorável à aprendizagem assistida por IA como o Aprend.AI pode incentivar discussões, trabalhos em grupo on-line e interações colaborativas que enriquecem a experiência educacional.

(v) Adaptação dinâmica de conteúdo: Com o uso de IA, é possível adaptar dinamicamente o conteúdo de aprendizagem com base no progresso do aluno, proporcionando um caminho personalizado para cada indivíduo – uma característica-chave das metodologias ativas.

## Sistemas de recomendação associados ao modelo de mudança

A grande virada de chave – e o que caracteriza o início da era da aprendizagem assistida por IA – são os sistemas de recomendação: tecnologias que utilizam algoritmos para analisar dados e fornecer sugestões personalizadas aos usuários com base nos seus interesses e preferências.

Esses sistemas são comuns em várias plataformas online e têm aplicações nos mais diferentes setores. Eles se baseiam em:

- Coleta de dados do usuário
- Análise de dados
- Filtragem colaborativa
- Filtragem baseada em conteúdo
- Personalização
- Atualização contínua
- Aplicações diversas

Podemos, ainda, acrescentar elementos dos sistemas de recomendação em nosso modelo de mudança de design ético para o uso de IAG, a fim de personalizar e direcionar a experiência de ensino sob a égide da ética e, desta forma, propiciar:

- (i) Coleta de dados éticos: Identificar métricas éticas e critérios relevantes para a avaliação do design ético em sistemas de recomendação. Questões como transparência do algoritmo, equidade nas sugestões e privacidade dos usuários devem ser cuidadosamente consideradas.
- (ii) Perfil do usuário ético: Desenvolver perfis éticos para os usuários, levando em conta suas preferências específicas em relação aos princípios éticos que desejamos enfatizar em nossa experiência educacional. Isso pode incluir a definição de prioridades éticas e valores específicos relacionados ao processo de ensino-aprendizagem.
- (iii) Algoritmos de recomendação ética: Desenvolver e implementar algoritmos de recomendação que considerem explicitamente os princípios éticos e critérios definidos. Isso pode envolver a adaptação de técnicas de filtragem colaborativa ou baseada em conteúdo para que possamos incorporar considerações éticas nas recomendações personalizadas.
- (iv) Aprendizado contínuo ético: Que o modelo permita a aprendizagem contínua sobre as preferências éticas dos usuários. À medida que os usuários interagem mais com o sistema, as recomendações devem ser ajustadas para refletir e respeitar melhor seus valores éticos evolutivos.
- (v) Transparência e explicabilidade: Garantir que nosso modelo seja transparente e forneça explicações claras sobre como as recomendações foram geradas com base nos critérios éticos estabelecidos, a fim de construir confiança e promover uma compreensão mais profunda por parte dos usuários sobre o processo de recomendação ética.

## **A mudança de paradigma: educação com IA**

Em um momento em que instituições acadêmicas globalmente se dedicam à elaboração de diretrizes para a implementação e uso de modelos de IA no âmbito da educação, a construção de um ambiente de

mudança torna-se crucial. Esse ambiente servirá de guia essencial para que a comunidade acadêmica – ela própria – estabeleça as bases para um processo educativo ativo, ágil e inclusivo, pautado, sobretudo, pela ética.

Nos anos de 2021 e 2022, a Unesco passou a gerar uma série de recomendações para o uso ético das IAs envolvendo os Estados-membros em uma colaboração internacional, para a difusão e uso de um conjunto abrangente de valores que passam pelo respeito, proteção e promoção dos direitos humanos, prosperidade ambiental, diversidade e inclusão, enfim, em total comprometimento com a dignidade humana. A aplicação dessas recomendações da Unesco (2023), para que a luz da ética incida sobre a forte presença das IAs no ensino superior, chega a abranger onze áreas de ação política para a operacionalização dos valores e princípios estabelecidos, para além da avaliação de impacto, governança e gestão éticos, alcançando áreas, como a do meio ambiente, gênero, cultura, saúde e bem-estar social.

A introdução das IAs generativas no processo educativo de professores, alunos e gestores, exige uma mudança assistida, reduzindo riscos iminentes e promovendo a colaboração mútua. Seguindo o princípio “nada sobre nós sem nós”, parafraseando Kofi Annan, essa mudança busca envolver todos os stakeholders de forma participativa e inclusiva. A revisão crítica das orientações publicadas nos últimos meses é essencial, dando prioridade àquelas que melhor se alinham com a cultura, história e costumes locais. Este esforço coletivo visa garantir que a implementação do uso de IAGs na educação seja não apenas tecnologicamente eficaz, mas também culturalmente sensível e socialmente responsável e ecologicamente sustentável. Este é apenas o começo de uma jornada em direção a um novo capítulo na evolução da educação concomitantemente à chegada de tecnologias disruptivas.

Em meio ao complexo tecido da vida acadêmica, emerge uma prática transformadora: a meta-mentoria. Este é o ato de mentorar o mentor, de guiar aqueles que, ao longo da jornada acadêmica, nos moldaram. É um mergulho profundo nas tecnologias digitais-instrucionais, uma jornada para preparar aqueles que nos ensinaram a decifrar as complexidades da mente artificial, a explorar possíveis alucinações criativas e a cocriar com modelos generativos ao longo da vida. Neste contexto, é crucial capacitar professores/mentores a comandar, como exímios *prompters*, a pavimentação do futuro do conhecimento.

Inspirados em visionários educacionais como Seymour Papert e Paulo Freire, entendemos que a aprendizagem de tecnologia é um agente catalisador de mudanças. A ideia de *Aprend.ai* (nosso micromundo educacional), torna-se o epicentro dessa transformação. Ela não apenas nos conecta à aprendizagem assistida por IA, mas cria um ambiente contro-

lado, conceito fundamental cunhado por Papert (1980) na sua linha construtivista, onde a meta-mentoria floresce. Esse ambiente controlado é um espaço seguro e estruturado, proporcionando segurança para explorar, experimentar, sobretudo, aprender. Nessa jornada, não apenas preparamos mentores, mas desenvolvemos lideranças para a nova fase educacional.

No que concerne à preparação de professores e educadores para o uso de IAGs, existem desafios a serem superados e lacunas a preencher: os programas educacionais deveriam estar orientados ao desenvolvimento de habilidades que lhes permitissem identificar e resolver problemas apoiados em técnicas, métodos e tecnologias computacionais.

Desenvolvido pela Unesco (2019), o Modelo de Competências Tecnológicas para Professores (*Technologies Competency Framework for Teachers* (ICTCFT)), aponta o que a docência deveria integrar as práticas profissionais, a fim de desenvolver o conhecimento e consciência críticos em seus alunos – em uma tentativa de germinar a literacia digital. Para Gabriel (2023), as habilidades do profissional na educação do futuro estariam amparadas pelos pilares: (1) da mentalidade digital (através de mudanças de paradigma); (2) do pensamento crítico e literacia digital e (3) da adaptabilidade (para poder transformar estratégias em realidade e sustentabilidade).

No âmbito educacional brasileiro, o atual desafio dos professores (em especial, os universitários) é o de preparar as novas gerações para conviver com as mudanças trazidas com a chegada das IAs. O letramento para uso de modelos de IAG na educação, assim como na pesquisa, sobressai-se entre as competências e habilidades digitais essenciais. Promulgada pelo Ministério da Ciência, Tecnologia e Inovações em 2021, a Estratégia Brasileira de Inteligência Artificial (EBIA) enfatiza a necessidade da aquisição e/ou desenvolvimento de um *hall* de competências digitais elementares para a educação e para a atualização da Base Nacional Comum Curricular (BNCC). A fim de incorporar elementos do pensamento computacional e programação, vislumbra-se desenvolvimento de programas abrangentes de alfabetização digital com a oferta de cursos sobre IA.

### **Meta-aprendizagem: metodologia conceitual na formação da docência superior**

Como metodologia conceitual, para pensarmos a formação de docentes do ensino superior baseada em ética e na regulamentação da IAG brasileira, enxergamos o modelo de meta-aprendizagem como um caminho possível nesse sentido. Proposta por Caio Vassão (2022), a meta-aprendizagem tem suas raízes influenciadas por *metadesign*, filosofia contemporânea, estudos de complexidade, antropologia estruturalista, ci-

bernética, pensamento crítico, fenomenologia, teoria do projeto e da arte, entre outras disciplinas. Na meta-aprendizagem, o educador desempenha um papel tanto de facilitador, quanto de aprendiz. Isso implica em uma transformação interna do educador, capacitando-o a acompanhar e entender o processo de aprendizagem daqueles que estão sob sua orientação.

A capacidade de guiar e liderar está intrinsecamente ligada à disposição do educador em evoluir e aprender de maneira autônoma. Segundo Gabriel (2023), na relação humano-tecnologia, o papel do educador se torna o vetor estrutural de mudança, conscientização e de responsabilidade. Mas, para entendermos a meta-aprendizagem, é necessário compreender que o *metadesign* “essa lógica que procura controlar a complexidade tornando-se um conjunto de objetos simples, de fácil compreensão” (Vassão, 2020, p. 15), sendo o estudo dos métodos que dialogam com sistemas.

Produzimos representações, percepções e conhecimentos a respeito da realidade que se organiza nessas camadas de complexidade. A abordagem generativa, algorítmica ou de procedimentos busca uma produção indireta por meio de “instruções” e conversões dessas entidades de forma visual, sonora, espacial, plástica, entre outras. Existem três tipos de sistemas que dialogam com o *metadesign*: (1) simples (deterministas com poucas variáveis), controlado e determinado; (2) complexo-desorganizado (pode ser tratado pela estatística e pela probabilidade) e (3) complexo-organizado (máquinas abstratas, por exemplo). A atividade do *metadesign* pode ser compreendida pela montagem e ajuste de máquinas abstratas a partir de um certo número de interações onde a ferramenta, mesmo amparada pelo digital, não é suficiente para lidar com a complexidade. “Nas áreas correlatas ao design de interação, existe uma demanda constante por critérios mais explícitos de projeto e avaliação chamada usabilidade” (Vassão, 2020, p.75).

De acordo com Shennan (2002), a cultura é uma estrutura com conexões transversais, complexa e interconectada com elementos que se entrelaçam de diversas maneiras. Compreender as relações culturais e suas ações no mundo contemporâneo, envolve organizar uma representação que construa um modelo explicativo do funcionamento dessa realidade – ou de uma parte dela. Portanto, ao desenvolver um modelo utilizando princípios do *metadesign*, fornecemos uma estrutura para entender as dinâmicas culturais em um contexto mais amplo.

Assim, a meta-aprendizagem (apoiada nos conceitos de *metadesign*) envolve uma progressão que contempla não somente a sequência de atitudes de um aprendiz em relação ao processo de aprendizagem, mas sobretudo a internalização de conceitos, a participação em atividades práticas, a reflexão sobre experiências e a capacidade de sintetizar e apli-

car o conhecimento adquirido de forma autônoma. A relação entre IAG e meta-aprendizagem pode ser entendida no contexto do desenvolvimento de sistemas educacionais mais adaptativos e personalizados.

Nesse sentido, o *Aprend.AI* corresponde à proposta de Vassão (2022) como um modelo de meta-aprendizagem, dentro do qual as relações se estabelecem em quatro momentos: (1) na experiência estética (E), (2) ontológica (O), (3) taxonômica (T) e (4) cosmológica (C) da aprendizagem – ocorrendo de forma orgânica e espontânea entre as diferentes experiências.

O processo de aprendizagem, quando abordado sob perspectivas estéticas, envolve a imersão em experiências inéditas, utilizando o corpo e a percepção de maneira individual ou coletiva. Este estágio prescinde de preocupações com nomenclaturas, denominações e vocabulário, promovendo uma abordagem visceral e concreta. Na esfera ontológica, a linguagem torna-se um instrumento crucial para compartilhar experiências, construir um vocabulário coletivo e estabelecer nomes e designações. A participação em debates coletivos contribui para a definição do que é considerado "real" em um dado contexto social e cultural. A taxonomia, por sua vez, demanda a sistematização do conhecimento, a criação de sistemas de pensamento e a produção de modelos sintéticos, incluindo expressões artísticas, promovendo a criatividade de forma estruturada e relevante. Na cosmologia, a ênfase recai sobre a transcendência de modelos estáticos, incentivando a visão do espaço entre sistemas e modelos.

Na aprendizagem em comunidade, aliada à segurança e autonomia para liderar processos de aprendizagem de maneira flexível e inclusiva, destaca-se reconhecer cada estágio anterior (estético, ontológico, taxonômico), e sua contribuição nas jornadas de aprendizagem é fundamental no percurso acadêmico. Nessa abordagem de Vassão, a meta-aprendizagem basicamente organiza em uma sequência de atitudes o que o aprendiz desenvolve em relação ao processo de aprendizagem.

O sistema educacional brasileiro, na sua maioria, com uma abordagem – ainda que essencialmente taxonômica, por vezes árida e desconectada da vida cotidiana – acaba por resultar em um ambiente acadêmico pouco estimulante e desmotivador para os alunos, pela predominância do texto e do livro como veículos centrais do conhecimento, reforçando uma visão sintética e reducionista.

Quando confrontados com temas abstratos, como matemática, geopolítica, economia ou astronomia, abordagens estéticas, engajando a memória dos aprendizes em experiências concretas relacionadas aos temas, permitiriam uma construção gradual de camadas conceituais mais abstratas, em contraponto à abordagem tradicionalmente taxonômica. Assim, a introdução de dimensões estéticas, ontológicas e cosmológicas,

utilizando aplicações de IAG, poderia criar uma experiência rica e significativa, de reconexão do conhecimento com o corpo, com as sensações, trazendo a promoção de uma visão plural e cosmológica que aceita ambiguidades e paradoxos, proporcionando uma experiência educacional mais envolvente e contextualizada.

## O aprendizado de IAGs na docência superior

Com base em tudo que foi visto até aqui e dentro de uma abordagem integrativa, holística e centrada no aprendiz, segue-se a apresentação de diferentes possibilidades da integração de IAGs, no ambiente controlado *AprenD.AI*, considerando a formação de professores para a docência superior amparada em um modelo de mudança de design ético. A abordagem associada à meta-aprendizagem e baseada em *metadesign*, assim se apresenta:

(i) Personalização e adaptação curricular: A IAG, aliada à meta-aprendizagem, pode ser usada para personalizar o currículo com base nas preferências e estilos de aprendizagem dos alunos, integrando a estética e ontologia propostas na metodologia. Para tanto, torna necessário considerar a diversidade de backgrounds culturais e experiências individuais, respeitando a autonomia do aprendiz.

(ii) Aprimoramento da experiência com a meta-aprendizagem: A integração de novas tecnologias de IAG à aula como, por exemplo, a realidade virtual e aumentada para proporcionar experiências estéticas conectando o aprendizado ao corpo e à percepção, podem ser estratégias para engajar o aluno ao conteúdo abordado de forma imersiva, bem como para explorar plataformas de colaboração online que facilitem discussões ontológicas, promovendo a construção de vocabulário compartilhado.

(iii) Sistematização do conhecimento e criatividade: A implantação de modelos de IAG com base na meta-aprendizagem, deve apoiar a taxonomia e a sistematização do conhecimento, permitindo a criação de modelos sintéticos e questões relativas à arte de forma colaborativa e inovadora, além de contribuir para a produção não apenas estética, como ética, de trabalhos criativos, respeitando direitos autorais e promovendo a originalidade.

(iv) Promoção da inclusão e da diversidade: Usar a IAG no contexto da meta-aprendizagem possibilita identificar possíveis vieses e garantir a inclusão de diferentes perspectivas nos materiais de aprendizagem, além de reforçar princípios cosmológicos de aprendizagem em comunidade. Essa incorporação possibilita que os professores elaborem estratégias adaptativas, como recursos personalizados, métodos de avaliação variados e atividades colaborativas que ampliam as perspectivas culturais.

(v) Monitoramento ético de desempenho (MED): Implementar sistemas de IAG dentro de um modelo de mudança de design ético apoiado pela meta-aprendizagem, para avaliação de desempenho, garante transparência, coíbe a discriminação e fornece autonomia aos alunos quanto à entrega de feedbacks personalizados e em respeito às diferentes trajetórias de aprendizagem.

(vi) Coleta ética de dados: Estabelecer políticas claras, com a inclusão da meta-aprendizagem, para a coleta e uso ético de dados dos alunos, garante a privacidade e a segurança, implementando práticas transparentes. A integração da meta-aprendizagem envolve o desenvolvimento de uma espécie de “algoritmocracia ética”, com a criação de comitês multidisciplinares para revisão constante das políticas vigentes. Educar os estudantes sobre o uso responsável de dados e proporcionar consentimento claro, são ações fundamentais para assegurar uma experiência educacional ílibada.

## Revitalização das práticas de avaliação

É fato que as IAGs já participam do processo educacional. Elas vieram para ficar e, se considerarmos que o método avaliativo de aprendizagem mais aplicado (ainda) nos nossos dias, data de 1930, é grande o contraste! A implementação de métodos como aprendizagem personalizada, sistemas inteligentes de tutoria, *chatbots* e realidade virtual está redefinindo o modo como os alunos são avaliados e, por esse motivo, é mister desenvolver competências na docência superior, de forma que professores e alunos possam estar integralmente nivelados – o que só será possível dentro de um modelo de mudança – que permita a execução de atividades práticas, elaboração de projetos interdisciplinares.

O *Aprend.AI* se vale de aparato tecnológico e de conhecimento especialista para mapear as atividades dos alunos em ambiente controlado, que substitui a aplicação de prova (no atual modelo top-down) por um processo que subverte a hierarquia da avaliação, oferecendo estímulos à construção de pensamento crítico pela curiosidade e investigação de temas desenvolvidos em conversas com IAGs, de modo a analisar suas investidas em diferentes contextos.

O modelo torna desnecessário especular se o aluno está utilizando IAG, porque ele é levado (natural e conscientemente) a interagir com essas inteligências na medida em que são disparadas requisições por parte dos professores-mentores. Com isso, ele é capaz de fornecer diagnósticos precisos, amparados pelo comportamento de seus usuários e que dão apoio para avaliações dinâmicas e o provimento de diretrizes que tornarão as práticas de ensino/aprendizagem mais assertivas.

Tudo dentro de uma abordagem ética, orientada aos objetivos e metas de conhecimento e que tende a ser cada vez mais personalizada, adaptativa-inclusiva e promotora de bem comum, corroborando a percepção global de que a análise e avaliação do progresso do aluno vem se tornando cada vez mais sofisticadas. É crucial que professores estejam conscientes de que seu desenvolvimento profissional depende do *life long learning*, ou seja, do seu aprendizado vitalício, observando, contudo, as Novas Práticas Avaliativas (NPAs), imbuídas não apenas da literacia digital, sobretudo

de valores dos quais as instituições de ensino (a despeito do grau), jamais devem abrir mão, em prol da integridade acadêmica e da retidão dos processos educacionais.

### Considerações finais

Em síntese, a adoção do modelo de mudança *Aprend.AI* revela-se como um design ético que reconfigura o tecido do ensino/aprendizagem, enriquecendo-o com uma visão plural e cosmológica que abraça ambiguidades e paradoxos inerentes ao processo educacional. Este modelo transcende as fronteiras convencionais, não apenas personalizando e adaptando currículos para professores e alunos, mas também elevando a experiência de aprendizagem através da integração de modelos de IA como tecnologias instrucionais e assistivas, impulsionando a consolidação da Aprendizagem Assistida por IA.

O design ético subjacente ao *Aprend.AI* não apenas se torna alicerçado na integridade acadêmica, mas também fornece um ambiente controlado, idealizado para permitir exploração, experimentação e aprendizado significativo. Aliado ao *metadesign* oferece estrutura, além de flexibilidade, adaptando-se dinamicamente às necessidades e evoluções do ambiente acadêmico.

A introdução da meta-mentoria redefine o papel do professor-mentor, capacitando-o não apenas a guiar, mas a ser guiado, criando assim um ciclo contínuo de aprendizado e desenvolvimento da classe. Da mesma forma, as Novas Práticas Avaliativas (NPAs) propõem uma abordagem inovadora para avaliar o progresso acadêmico, destacando não apenas a aquisição de conhecimento, mas também o desenvolvimento de habilidades críticas e adaptativas.

Portanto, ao consolidar a aprendizagem assistida por IA como uma prática inovadora e ética, o *Aprend.AI* não só responde às demandas contemporâneas, mas também fortalece o ambiente acadêmico, capacitando gestores, professores e alunos a enfrentar os desafios tecnológicos presentes e futuros com resiliência e discernimento. Esta abordagem abraça a metamorfose constante do conhecimento, estabelecendo um paradigma educacional adaptável, ético e sustentável.

### Sistema de recomendações – *aprend.ai*

Lembremo-nos de que nenhuma técnica de comunicação, do telefone à Internet, traz por si mesma a compreensão.[...] Ensinar a compreensão entre as pessoas como condição e garantia da solidariedade intelectual e moral da humanidade. (Morin, 2002, p. 93)

Para a composição desse documento, extraímos deste capítulo – Aprend.AI: um modelo ético das IAGs para o ensino superior – um conjunto de recomendações acerca do que se deve evitar, alertar e incentivar. Trata-se de um framework essencial para a preservação da integridade acadêmica, sobretudo ao apoio à dinâmica de transição para a Era da Aprendizagem Assistida por IA, como se segue:

## 1. Benefícios e desafios da mudança de paradigma

### Evitar

- **Subutilização de práticas inclusivas:** Evitar a falta de adoção de práticas educacionais que não aproveitem plenamente as capacidades das IAGs para criar um ambiente de aprendizagem ativo, ágil e inclusivo.
- **Desconsideração de valores éticos e fundamentais:** Evitar a implementação de tecnologias de IA que ignorem os valores de respeito, diversidade e sustentabilidade, fundamentais para uma educação ética e inclusiva.

### Alertar

- **Necessidade de uma mudança assistida e consciente:** Alertar sobre a importância de uma transição cuidadosamente planejada e assistida para a implementação de IAGs, de forma a minimizar riscos e garantir uma adaptação eficaz e colaborativa entre todos os envolvidos.
- **Importância da revisão crítica e adaptação cultural:** Alertar para a necessidade de uma revisão crítica das práticas e orientações existentes, garantindo que a implementação de IAGs esteja alinhada com a cultura, história e costumes locais, para uma integração culturalmente sensível e socialmente responsável.
- **Preparação adequada dos educadores:** Alertar para o desafio de preparar os educadores para integrar eficazmente as IAGs em suas práticas pedagógicas, destacando a necessidade de programas educacionais focados no desenvolvimento de habilidades relevantes para a era digital.

### Incentivar

- **Promoção da educação ativa e inclusiva:** Incentivar o uso de IAGs para enriquecer o processo educativo, tornando-o mais ativo, ágil e inclusivo, melhorando a experiência de aprendizagem de professores, alunos e gestores.
- **Adoção de uma ética baseada em valores fundamentais:** Incentivar a implementação de IAGs que se baseiem nas recomendações éticas da UNESCO, promovendo valores como respeito, diversidade e sustentabilidade em todas as práticas educacionais.
- **Estímulo à colaboração participativa:** Incentivar a adoção do princípio de "nada sobre nós sem nós", especialmente em comunidades historicamente marginalizadas, para fomentar a colaboração participativa de todos os stakeholders na implementação das IAGs, assegurando uma abordagem verdadeiramente inclusiva.

## 2. Aproximação do conceito de meta-aprendizagem

### Evitar

- **Simplificação excessiva do processo de aprendizagem:** Evitar a redução do processo de aprendizagem a modelos simplistas que não capturam a complexidade do conhecimento humano e suas nuances interdisciplinares.

### Alertar

- **Complexidade da integração de IAs:** Alertar para os desafios e a complexidade de integrar eficazmente as IAs na meta-aprendizagem, ressaltando a necessidade de abordagens cuidadosas que respeitem as experiências estéticas e aprofundem a cosmovisão dos estudantes.

### Incentivar

- **Adoção da meta-aprendizagem para formação ética:** Incentivar educadores a explorar a meta-aprendizagem e o metadesign para uma formação ética abrangente no ensino superior, promovendo a capacidade de liderar e aprender de maneira autônoma.
- **Integração holística das IAs:** Incentivar a incorporação de IAs de maneira holística e interdisciplinar na educação, melhorando as experiências de aprendizagem através da ampliação da cosmovisão e facilitação de taxonomias com algoritmos.
- **Enriquecimento da educação com abordagens estéticas e ontológicas:** Incentivar o uso de tecnologias como IA, IAGs, realidade virtual e realidade aumentada para criar experiências educacionais ricas e envolventes, conectando o conhecimento ao corpo e às sensações e promovendo uma educação mais contextualizada.

## 3. A importância do aprendizado contínuo sobre IAs na docência superior

### Evitar

- **Desconsiderar a evolução tecnológica:** Evitar ignorar a rápida evolução das IAs e seu impacto potencial sobre a educação, que pode resultar em currículos desatualizados e práticas de ensino ineficazes.
- **Negligenciar o desenvolvimento de habilidades técnicas e emocionais:** Evitar a falta de ênfase no desenvolvimento de habilidades técnicas e emocionais essenciais para a navegação eficaz na era digital, limitando a capacidade dos educadores de se adaptarem às novas demandas educacionais.

### Alertar

- **Necessidade de adaptação contínua:** Alertar sobre a importância da adaptação contínua às inovações tecnológicas e colaboração entre educadores para manter as práticas pedagógicas relevantes e eficazes.
- **Importância da ética no redesign educacional:** Alertar para a necessidade de incorporar princípios éticos no processo de mudança, assegurando que a introdução das IAs na educação seja conduzida de maneira responsável e inclusiva.

### Incentivar

- **Promoção do *Lifelong Learning*:** Incentivar os educadores a adotarem uma postura de aprendizado contínuo, reconhecendo o papel crucial que desempenham na era das IAs e a importância de se manterem atualizados sobre essas tecnologias.
- **Fomento à exploração e colaboração:** Incentivar as instituições de ensino superior a criar ambientes que promovam a exploração de novas inovações e a colaboração entre educadores, facilitando a adaptação às mudanças tecnológicas.
- **Utilização de modelos de mudança para qualificação:** Incentivar o uso de modelos de mudança para melhor qualificar os professores, preparando-os para atuar como mentores proficientes, curadores de conteúdo e gestores de tecnologia instrucional.
- **Desenvolvimento de competências multidisciplinares:** Incentivar o desenvolvimento de uma ampla gama de competências, incluindo habilidades técnicas e emocionais, para liderar a produção acadêmica de alto nível e garantir uma aplicabilidade efetiva no contexto das IAs.
- **Integração da ética no processo educacional:** Incentivar a incorporação de princípios éticos no redesign do processo educacional, promovendo uma transição sensível e alinhada aos valores fundamentais da comunidade acadêmica.

## O DODECÁLOGO – Síntese das recomendações elaboradas sob o Aprend.AI para a adequação das IES ao sistema de Aprendizagem Assistida por IA.

Diante das transformações educacionais impulsionadas pela Aprendizagem Assistida por IA (AAIA), derivadas do modelo ético Aprend.ai, sintetizamos as recomendações-chave na forma de um dodecálogo que surge como resposta às demandas de uma educação personalizada, ética e adaptativa.

### Evitar

1. **Ignorar a diversidade cultural e de aprendizagem:** Evitar a padronização excessiva que desconsidera as diferenças individuais e culturais dos alunos, comprometendo a inclusão e a personalização do ensino.
2. **Subestimar a privacidade e a segurança dos dados:** Evitar o uso negligente de tecnologias que possam comprometer a privacidade e a segurança dos dados dos alunos, respeitando sempre as normativas e as expectativas de confidencialidade.
3. **Depender unicamente de soluções tecnológicas para avaliação:** Evitar a dependência exclusiva em sistemas de IA para avaliação e feedback, reconhecendo a importância da interação humana e do julgamento pedagógico.
4. **Desconsiderar o impacto emocional da IA na aprendizagem:** Evitar o uso de IA sem considerar seu impacto emocional e social nos alunos, garantindo que a tecnologia sirva como suporte, e não como substituto das relações humanas

## Alertar

- 1. Transparência na utilização de IA:** Alertar sobre a necessidade de transparência nos algoritmos de IA, permitindo que alunos e professores compreendam como as decisões são tomadas pelas ferramentas que utilizam.
- 2. Riscos de vieses e discriminação:** Alertar para os riscos de vieses inerentes aos sistemas de IA, trabalhando ativamente para mitigar a discriminação e promover a equidade na educação.
- 3. Necessidade de formação contínua sobre ética em IA:** Alertar sobre a importância da formação contínua de professores e gestores em questões éticas relacionadas ao uso de IA, assegurando uma integração consciente e responsável.
- 4. Desafios na adaptação curricular:** Alertar para os desafios de integrar IA de forma significativa no currículo, garantindo que a tecnologia complemente e enriqueça o processo de aprendizagem sem sobrecarregar alunos ou professores.

## Incentivar

- 1. Promoção de práticas pedagógicas inovadoras:** Incentivar o uso criativo e inovador de IA para enriquecer as práticas pedagógicas, apoiando a personalização da aprendizagem e a exploração de novas metodologias de ensino.
- 2. Desenvolvimento de habilidades digitais e críticas:** Incentivar o desenvolvimento de habilidades digitais e críticas entre alunos e professores, preparando-os para interagir de forma eficaz e ética com a tecnologia.
- 3. Colaboração interdisciplinar no design de IA:** Incentivar a colaboração entre educadores, técnicos e estudantes no design e implementação de soluções de IA, promovendo abordagens interdisciplinares e participativas.
- 4. Fortalecimento da literacia em IA:** Incentivar programas que aumentem a literacia em IA, tanto para alunos quanto para professores, garantindo que a comunidade educativa esteja preparada para utilizar, questionar e contribuir para o desenvolvimento de tecnologias éticas e inclusivas.

Esperamos que nossas recomendações surtam efeito positivo na prática docente, sobretudo por propor que tais mudanças aconteçam de forma colaborativa-inclusiva em um ambiente controlado apropriado à incorporação assistida de IAGs, à medida que auxiliará a superar resistências e redefinir o papel do docente durante essa transformação crucial para o avanço educacional.

## Referências

ABEND, Gabriel. *The moral background: an inquiry into the history of business ethics*. Princeton, NJ: Princeton University Press, 2014. Disponível em: [https://www.researchgate.net/publication/275154538\\_The\\_Moral\\_Background\\_An\\_Inquiry\\_into\\_the\\_History\\_of\\_Business\\_Ethics](https://www.researchgate.net/publication/275154538_The_Moral_Background_An_Inquiry_into_the_History_of_Business_Ethics). Acesso em: 2 jan., 2024.

ALAN, Sokal. Transgressing the boundaries: towards a transformative hermeneutics of quantum gravity. *Social Text*, v. 46/47, p. 217-252, 1996. Disponível em: [https://physics.nyu.edu/sokal/transgress\\_v2/transgress\\_v2\\_singlefile.html](https://physics.nyu.edu/sokal/transgress_v2/transgress_v2_singlefile.html). Acesso em: 24 jan., 2024.

ALEXANDER, Jonathan. Student's right to write. Inside Higher, Ed. 2023. Disponível em: <https://www.insidehighered.com/opinion/views/2023/11/22/students-have-right-write-ai-era-opinion>. Acesso em: 15 jan., 2024.

ANNAN, Kofi. Secretary-General Kofi Annan's message on the International Day of Disabled Persons, observed on 3 December 2004. Disponível em: <https://press.un.org/en/2004/sgsm9628.doc.htm>. Acesso em: 21 jan., 2024.

ARAYA, Roberto. Teacher training, mentoring or performance support systems? *Advances in Intelligent Systems and Computing*. Centro de Investigación Avanzada en Educación, Universidad de Chile, Santiago, 2018. Disponível em: [https://repositorio.uchile.cl/bitstream/handle/2250/169660/Teacher\\_training.pdf](https://repositorio.uchile.cl/bitstream/handle/2250/169660/Teacher_training.pdf). Acesso em: 21 jan., 2024.

BALL, Philip. Computer conference welcomes gobbledegook paper. *Nature*, London, v. 434, p. 946 21 April, 2005. Disponível em: <https://rdcu.be/dwsHL>. Acesso em: 5 jan., 2024.

CHATGPT USED BY TEACHERS more than students, new survey from Walton Family Foundation finds. *Walton Family Foundation Education*, March 1, 2023. Disponível em: <https://www.waltonfamilyfoundation.org/chatgpt-used-by-teachers-more-than-students-new-survey-from-walton-family-foundation-finds>. Acesso: 10 jan., 2024.

DELEUZE, Gilles, GUATTARI, Félix. *Mil Platôs: capitalismo e esquizofrenia*. Tradução: Aurélio Guerra Neto, et. al., Rio de Janeiro: Editora 34, 1995.

FERREIRA, Sulane C. Apartheid digital em tempos de educação remota: Atualizações do racismo brasileiro. *Interfaces Científicas – Humanas e Sociais*, Aracaju, v. 10, n. 1, p. 11-24, 2020.

FERRY, Luc. *A inovação destruidora: ensaio sobre a lógica das sociedades modernas*. Rio de Janeiro: Objetiva, 2015.

GABRIEL, Martha. *Educação na era digital: conceitos, estratégias e habilidades*. São Paulo: Atlas, 2023.

GATTI, Daniel C.; FONSECA, Tina. O design ético de modelo de mudança para o uso de IAG como ferramenta de ensino. Projeto de pesquisa apresentado no 4º Congresso de Inteligência Artificial da PUC-SP. São Paulo: 7-8 nov. 2023. Disponível em: <https://www.youtube.com/@TIDDPUCSP>. Acesso em: 10 jan., 2024.

GREENE, Daniel; HOFFMAN, Anna Lauren; STARK, Luke. Better, nicer, clearer, fairer: A critical assessment of the movement for ethical artificial intelligence and machine learning. In: BUI, Tung (ed.), *52nd Hawaii International Conference on System Sciences, HICSS 2019*, Grand Wailea, Maui, HI, 8-11 Jan. 2019. Disponível em: <https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/849782a6-06bf-4ce8-9144-a93de4455d1c/content>. Acesso em: 10 jan., 2024.

LABBÉ, Cyril. Ike Antkare one of the great stars in the scientific firmament. *International Society for Scientometrics and Informetrics Newsletter*, 2010, ed. 6, vol. 2, p. 48-52. Disponível em <https://hal.science/hal-00713564/document>. Acesso em: 20 jan., 2024.

LOVE, Ruth B. De uma administradora escolar. In: ADLER, M. (Org.) *A proposta Paidéia: um manifesto educacional*. Campinas: Cedet, 2021. p. 87-88.

NIELSEN, Jacob. 10 Usability heuristics for user interface design. Nielsen Norman Group, 1994. Disponível em: <https://www.nngroup.com/articles/ten-usability-heuristics>. Acesso em: 21 jan., 2024.

MORIN, Edgar. *Os sete saberes necessários à educação do futuro*. São Paulo, Brasília: Cortez / Unesco, 2002.

PAPERT, Seymour. *Mindstorms: children, computers and powerful ideas*. New York: Nasic Books, 1980.

PEDRO, Francesc, *et al.* Artificial intelligence in education: Challenges and opportunities for sustainable development. Paris: UNESCO (= Working Papers on Education Policy, 7). Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000366994>. Acesso em: 10 jan., 2024.

SANTAELLA, Lucia. *Há como deter a invasão do CHATGPT?* São Paulo: Estação das Letras e Cores, 2023.

SHENNAN, Stephen. *Genes, memes and human history: Darwinian archaeology and cultural evolution*. London: Thames & Hudson, 2002.

SHILTON, Katie. Engaging values despite neutrality. *Science, Technology & Human Values*, London, v. 43, n. 2, 2018. Disponível em: <https://doi.org/10.1177/0162243917714869>. Acesso em: 22 jan. 2024.

VAN NOORDEN, Richard. The ethical questions that haunt facial-recognition research. *Nature*, London, v. 587, n.7834, p.:354-358, nov. 2020. Disponível em: [go.gale.com/apps/doc/A650829229/AONE?u=anon~7abbd4d2&sid=googleScholar&xid=2277a752](https://go.gale.com/apps/doc/A650829229/AONE?u=anon~7abbd4d2&sid=googleScholar&xid=2277a752). Acesso em: 22 jan., 2024.

UNESCO IESALC (Instituto Internacional de la UNESCO para la Educación Superior en América Latina y el Caribe). ChatGPT, artificial intelligence and higher education. *World Education Blog*, 2023. Disponível em: <https://www.iesalc.unesco.org/en/2023/04/14/chatgpt-and-artificial-intelligence-in-higher-education-quick-start-guide-and-interactive-seminar/>. Acesso em: 8 jan., 2024.

UNESCO. *Artificial Intelligence in Education: Challenges and opportunities for sustainable development*. Education Sector, 2019. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000366994>. Acesso em: 10 jan., 2024.

UNESCO. *Para aproveitar a era da Inteligência Artificial na educação superior: um guia às partes interessadas do ensino superior*, 2023. Disponível em: <https://en.unesco.org/open-access/terms-use-ccbysa-port>. Acesso em: 10 jan., 2024.

VASSÃO, Caio Adorno. *Metadesign: ferramentas, estratégias e ética para a complexidade*. São Paulo: Blucher, 2020.

VASSÃO, Caio Adorno. Metadesign e meta-aprendizagem: da estética ao cosmo. *Medium*, mar. 14, 2022. Disponível em: <https://caiovassao.medium.com/metadesign-e-meta-aprendizagem-da-est%C3%A9tica-ao-cosmo-1f35b7b9a92b>. Acesso em: 8 jan., 2024.

VASSÃO, Caio Adorno; FREITAS, Julio César; MARIN, Túlio C.T. Infraestrutura em computação pervasiva para suporte à pesquisa acadêmica colaborativa. In: *Anais do II Congresso Internacional de Design da Informacao*. São Paulo: Senac, 2005.

WICKRAMASINGHE, Bimali. Human-computer interaction – principles, evaluation and universal design principle. *Medium*, 2020. Disponível em: <https://bimalics.medium.com/human-computer-interaction-principles-evaluation-and-universal-design-principle-3687123b5b2a>. Acesso em: 15 jan., 2024.

# Guia Ético para a Inteligência Artificial Generativa no Ensino Superior

Diego Franco<sup>1</sup>

Luís Eduardo Viegas<sup>2</sup>

Anderson Röhe<sup>3</sup>

dx.doi.org/  
10.23925/1984-3585.2023i28p108-117

Licensed under  
[CC BY 4.0](#)

**Resumo:** Este guia visa orientar não só professores, alunos, administradores, quanto a própria instituição de ensino superior sobre o uso ético e responsável da Inteligência Artificial Generativa – (IAG) na Educação. As regras estão distribuídas em quatro categorias principais: 1. ética da instituição de ensino superior, 2. ética dos professores no uso da IAG com os alunos, 3. ética dos alunos e pesquisadores em redação e pesquisas, 4. ética na formação de grupos e comitês dentro da comunidade acadêmica.

**Palavras-chave:** educação; ensino superior; ética; Inteligência Artificial Generativa; inteligência artificial responsável

---

<sup>1</sup> Doutorando em Tecnologias da Inteligência e Design Digital pela PUC-SP, mestre em Tecnologias da Inteligência e Design Digital pela PUC-SP, especialista em Gestão Estratégica de Marketing pela FDC. Atualmente, sua pesquisa mergulha na atuação da IA Generativa, explorando a intersecção entre tecnologia, criatividade e cultura. Também é diretor de criação e designer multidisciplinar atuando em empresas e agências de publicidade no Brasil. Orcid: <https://orcid.org/0009-0005-7379-3894>.

<sup>2</sup> Doutorando em Tecnologias da Inteligência e Design Digital (TIDD) pela PUC-SP, Mestre em Administração pelo IBMEC e Graduado em Informática pela UERJ. Pesquisador em IA Responsável pela Cátedra Oscar Sala do Instituto de Estudos Avançados da USP, em Tecnologias Emergentes pelo Grupo de Pesquisa Transformação Digital e Sociedade da PUC-SP, e pelo Observatório do Futuro do Trabalho. Orcid: <https://orcid.org/0009-0003-5334-021X>.

<sup>3</sup> Doutorando em Tecnologias da Inteligência e Design Digital (TIDD) pela PUC SP. Mestre em Políticas Internacionais pela PUC Rio. Advogado graduado pela UFF. Especialista em Direito Digital pela UERJ e ITS Rio. Membro da Comissão Especial de Privacidade, Proteção de Dados e Inteligência Artificial da OAB-SP. Pesquisador em IA Responsável pela Cátedra Oscar Sala do Instituto de Estudos Avançados da USP. Orcid: <https://orcid.org/0000-0002-3104-6365>.

## **Ethical guide for the use of Generative Artificial Intelligence in higher education**

**Abstract:** This document aims at guiding students, teachers, and higher education institutions personnel with respect to best practices for the ethical and responsible use of Generative Artificial Intelligence in education. The rules are organized into four groups: (1) ethics for higher education institutions, (2) ethics for teachers in using Generative Artificial Intelligence in class, (3) ethics for students and researchers in writing and research, and (4) ethics in forming groups and committees in the academic community.

**Keywords:** education; ethics; Generative Artificial Intelligence; higher education; responsible artificial intelligence

## I. **Ética da instituição de ensino superior**

### 1.1 *Quanto às medidas de prevenção e responsabilidade*

É fundamental que as instituições de ensino superior tenham suas próprias políticas de prevenção e procedimentos específicos de contenção de danos, avaliação de impacto e mitigação de riscos a fim de assegurar o bem-estar, a segurança e os direitos de todos os envolvidos, sobretudo com relação à privacidade e proteção de dados, conforme as leis nacionais vigentes e as boas práticas internacionais quanto ao uso ético e responsável da IAG na Educação.

Para tanto, as instituições devem constituir comitês de crise/emergência para assegurar que resultados inesperados ou situações críticas sejam tratadas com rapidez, mitigando assim o dano ao máximo que possível, seguindo métodos pré-estabelecidos e a devida representatividade.

Já quanto à responsabilidade, esta deve ser compartilhada, na proporção da contribuição de cada um para a produção de eventual dano oriundo de abuso ou mal uso da IAG, cabendo não só ao professor seguir as políticas e procedimentos da instituição, mas a comunidade acadêmica como um todo.

### 1.2 *Quanto à privacidade e proteção de dados*

Implementar medidas de segurança e salvaguardas de transparência, necessidade, proporcionalidade, razoabilidade e consentimento para o uso de dados dos usuários contra abusos, acessos indevidos, não autorizados e/ou eventuais vazamentos de informação sensível e/ou sigilosa, conforme as diretivas mais relevantes quanto à privacidade e proteção de dados, tais como o Regulamento Geral de Proteção de Dados – GDPR europeu e a Lei Geral de Proteção de Dados – LGPD no Brasil, garantindo privacidade e transparência dentro de sala de aula.

O propósito é de que os professores não devam ser os únicos responsáveis por eventuais prejuízos e/ou causação de danos, mas a instituição como um todo, na proporção de sua contribuição. Caberia ao professor, individualmente, zelar e seguir as regras da respectiva instituição.

### 1.3 *Quanto à equidade*

Monitorar regularmente o uso das ferramentas de IAG por alunos, pesquisadores e professores para identificar, substituir ou mitigar quaisquer vieses que possam levar a tratamentos injustos ou discriminatórios, assegurando a igualdade de oportunidades para todos, independentemente de sua origem, gênero ou identidade étnico-racial.

#### 1.4 *Quanto à fiscalização*

As instituições, como parte de suas estruturas de governança para o uso ético e responsável da IAG, devem implementar políticas e procedimentos para detecção de plágio, evitando a indesejada violação de direitos autorais e propriedade intelectual. Uma vez que não se deve ignorar o potencial da IAG de produzir conteúdo falso ou que possa vir a ser considerado plágio.

A falta de fiscalização pode resultar em violações éticas e prejudicar a reputação não só do pesquisador, da pesquisa em si, mas da própria instituição. Cabe, então, aos pesquisadores, subsidiariamente à instituição de ensino, adotar procedimentos e ferramentas disponibilizados por aquela, a fim de realizar suas produções de acordo com a necessária conduta ética.

## 2. **Ética dos professores no uso da IAG com os alunos**

#### 2.1 *Quanto à explicabilidade e transparência*

Descrever sempre prévia e detalhadamente os critérios de escolha, as ferramentas de IAG utilizadas para gerar determinado conteúdo (texto, áudio e imagem) e como serão aplicadas nas atividades educacionais, incluindo quais são os dados, a origem de suas fontes, o propósito de sua utilização e para quais finalidades. Posteriormente, esclarecer quais serão os benefícios e as limitações do uso de IAG dentro daquele contexto específico, ajudando os alunos a compreenderem melhor como essa tecnologia poderá afetar sua aprendizagem. O intuito é de explicar e justificar a tomada de decisão, a fim de gerenciar o problema da falta de transparência e accountability ainda presentes em sistemas de IA.

#### 2.2 *Quanto ao consentimento informado e esclarecido*

Obter o consentimento informado e esclarecido dos alunos – isto é, decorrente da sua manifestação livre, explícita e inequívoca – antes de administrar seus dados em sistemas de IAG, assegurando que estejam cientes de como, onde e até quando suas informações serão utilizadas e armazenadas, fornecendo alternativas para que optem ou não por participar, se assim desejarem, isentos de eventuais retaliações, penalidades e/ou quaisquer outros prejuízos à sua aprendizagem.

Sempre que possível, evitar utilizar dados pessoais dos alunos em sistemas de IAG. Em caso de estrita necessidade de utilizar dados de alunos, assegurar que os envolvidos tenham ciência prévia e tenham a opção de concordar ou não com a sua utilização. Se a docência achar que o uso de IAG poderá gerar benefícios adicionais para o processo de aprendiza-

gem e/ou desenvolvimento dos discentes, a possibilidade de ocorrer prejuízos deve ser esclarecida em linguagem aberta e de fácil compreensão. Ainda assim, a vontade do discente em ceder ou não seus dados deverá prevalecer e ser respeitada. O propósito é evitar a banalização no uso de dados pessoais de modo desnecessário, desarrazoado e desproporcional.

### *2.3 Quanto à inclusão e acessibilidade*

Como parte de uma instituição que presta contas à sociedade, é essencial que as ferramentas de IA, se usadas, sejam acessíveis a todos os alunos, independentemente de diferenças socioeconômicas e discriminações algorítmicas, adaptando a tecnologia para responder bem às necessidades de aprendizagem que consigam promover o justo equilíbrio de oportunidades nos ambientes educacionais.

### *2.4 Quanto à autonomia e o pensamento crítico do aluno*

O desenvolvimento acadêmico dos alunos é o propósito elementar da instituição. Portanto, é necessário promover o desenvolvimento e a independência intelectual dos alunos ao incentivá-los a usar a IAG como uma ferramenta de apoio, aliada ao aprendizado, e não como substituta do esforço pessoal, bom senso e pensamento crítico, encorajando o trabalho analítico que sempre considere os possíveis impactos da IAG na própria aprendizagem.

## **3. Ética dos alunos e pesquisadores em redação e pesquisas**

### *3.1 Quanto à transparência na redação e pesquisa*

Sempre que o aluno ou pesquisador utilizar IAG para redigir, criar conteúdo ou simplesmente obter auxílio em suas pesquisas é fundamental explicitar o grau do uso da tecnologia. Isso inclui especificar quais partes do trabalho foram geradas por IA e quais foram elaboradas ou editadas sem sua participação. Evitar que o aluno apresente o conteúdo gerado como se fosse exclusivamente seu e/ou apenas de autoria humana. A omissão do uso da IAG em trabalhos acadêmicos pode comprometer a lisura, transparência e integridade dos processos decisórios com relação a critérios determinantes para a obtenção de distinção ou excelência acadêmica, como inovação, ineditismo e originalidade/autenticidade.

### *3.2 Quanto à verificação de fontes*

Antes de utilizar IAG na produção acadêmica, é preciso verificar a consistência dos conteúdos por esta gerados. Isso significa revisar cuidadosamente as fontes utilizadas, preferindo dados precisos e confiáveis

para confirmar a veracidade dos conteúdos. Portanto, evitar a presunção de que o conteúdo produzido por IA é livre de erros ou imprecisões. O otimismo cego sobre a precisão e a confiabilidade da tecnologia pode levar à equívocos e, em um grau maior, à disseminação de desinformação (informações imprecisas, incorretas ou enganosas).

### *3.3 Quanto à garantia de originalidade*

Usar a tecnologia tanto como ferramenta para inspirar a criatividade, quanto suporte no desenvolvimento e promoção de ideias próprias, de modo a não permitir que a IAG substitua o processo criativo, não empurrando para a tecnologia a tarefa de encontrar o substrato científico de suas produções.

### *3.4 Quanto à prevenção ao plágio*

O aluno ou pesquisador deve se informar previamente, seguindo os regramentos de sua instituição de ensino para a detecção de plágio, evitando a violação não só do ineditismo da obra, quanto a de direitos autorais de terceiros. Não se deve ignorar o potencial da IAG de produzir conteúdo que possa vir a ser considerado plágio.

### *3.5 Quanto ao processo contínuo de aprendizado*

Cabe ao aluno ou pesquisador se atualizar e se manter sempre informado sobre o avanço de marcos regulatórios e legislações específicas sobre usos e implicações da IAG na educação. Seja participando, seja incentivando a comunidade a participar de eventos e seminários que se debruçam sobre o tema. A estagnação no aprendizado pode levar ao uso inadequado ou obsoleto da tecnologia.

### *3.6 Quanto ao diálogo multidisciplinar*

Não se esquivar do diálogo multidisciplinar sobre as implicações éticas da IAG. Ignorar sua complexidade pode resultar em uma compreensão limitada, muitas vezes enviesada, sobre a responsabilidade ética associada ao seu uso. Promover discussões francas e abertas, conectando pesquisadores de outras áreas sobre os seus desafios e oportunidades. Isso inclui explorar, independente da área de pesquisa, questões de viés, privacidade e o impacto futuro da IA na sociedade.

## **4. Ética na formação de grupos e comitês**

### *4.1 Quanto à inclusão e diversidade*

Garantir uma composição diversificada e inclusiva de grupos de gerenciamento de crises e comitês de emergência, representando a pluralidade de disciplinas, perspectivas e experiências adquiridas é um outro tema fundamental. Isso implica a inclusão de membros de diferentes departamentos acadêmicos, especialistas em tecnologia, representantes

estudantis e membros do corpo administrativo, sobretudo os que tenham domínio sobre aspectos legais, éticos e organizacionais inerentes ao uso da tecnologia na respectiva instituição. Não limitar a formação do grupo a um único campo de estudo ou perspectiva, o que poderia resultar em uma compreensão parcial e restrita dos impactos da tecnologia. Estimular a pluralidade étnica, racial, de gênero, social, etária e religiosa como parte da formação desses grupos e comitês.

#### *4.2 Quanto à clareza e lisura nas operações*

As operações de grupos e comitês devem ser mantidas em caráter aberto e transparente, divulgando regularmente suas metas, objetivos e resultados alcançados a fim de emitir recomendações periódicas à comunidade acadêmica. Isso promove tanto a credibilidade, quanto a legitimidade, já que permite que todos estejam informados e, de alguma forma, envolvidos nos processos de decisão. Não conduzir as atividades em segredo ou sem uma comunicação verdadeiramente clara com os interessados, pois isso pode gerar desconfiança e especulações temerárias.

#### *4.3 Quanto ao diálogo contínuo*

O diálogo institucional deve ser promovido de maneira aberta, incluindo a realização de fóruns, seminários e consultas públicas. Isso permite não apenas o intercâmbio de ideias, angústias e inquietações, mas a revelação das expectativas sobre o uso da IAG na educação e sociedade como um todo. Não ignorar ou minimizar as preocupações e contribuições de pesquisadores, instituições e think tanks em função de aparente prestígio de umas em detrimento de outras. A falta de engajamento realmente democrático pode levar a decisões que não representem a pluralidade da instituição e resistências às iniciativas propostas.

#### *4.4 Quanto à educação e capacitação*

Devem ser desenvolvidos programas de educação e capacitação para membros, assim como não membros do comitê, expandindo os horizontes de informação e pesquisa e visando aumentar a compreensão sobre a IAG e seus impactos éticos, legais e sócio técnicos. Não deve ser assumido que todos os membros do grupo estejam no mesmo nível de compreensão, conhecimento ou interesse pelo tema. Fortalecer, portanto, a interlocução independentemente do nível de formação, uma vez que a ausência dessas trocas de conhecimentos pode inibir o aperfeiçoamento educacional e comprometer o próprio propósito do grupo ou comitê.

#### *4.5 Quanto à análise crítica e reflexiva*

Não devem ser adotadas abordagens exclusivamente tecnocráticas e/ou que desconsiderem as complexas implicações éticas e morais sobre o uso de ferramentas tecnológicas. Deve ser encorajada uma análise

holística que se ocupe dos reais impactos da IAG, considerando tanto os benefícios quanto os riscos em potencial. O grupo deve ser plural o suficiente para ponderar sobre questões éticas, legais e sociotécnicas, formulando recomendações que considerem o benefício de toda a comunidade.

#### 4.6 Quanto à interlocução e às parcerias institucionais

Grupos e comitês devem manter contato ativo com outros, reforçando parcerias institucionais, ampliando o diálogo entre os vários grupos de interesses e ajustando suas estratégias, quando necessário. A tecnologia está em constante evolução e as abordagens precisam ser flexíveis para se adaptarem a novos momentos e contextos. Não se apegue, portanto, a conceitos ou políticas inflexíveis. Resistências puramente personalistas podem impedir o grupo ou comitê de responder eficazmente aos desafios e oportunidades apresentados pela IAG.

O compromisso ético no ensino superior, diante dos desafios impostos pela IAG na educação, transcende a mera formalidade institucional. Ele é fundamental para a preservação de um modelo educacional que atenda efetivamente às demandas da sociedade. Este manual, portanto, não é apenas um guia para o uso ético e responsável da IAG, promovendo um ambiente educacional mais diverso e inclusivo, mas também se arrisca a definir diretrizes claras que preservem a integridade acadêmica e fomentem a inovação consciente. Na prática, um manual de ética atua como uma bússola em um ambiente de incertezas, assegurando que as políticas adotadas sejam justas e abrangentes; entretanto, sua existência sempre enfatiza a necessidade de revisão e atualização para adaptá-las a novos contextos e descobertas.

## GUIA PRÁTICO PARA O USO DA IAG



Estabelecer e esclarecer a política de uso de IA para toda a população da instituição, assegurando que a política seja aderente ao propósito da instituição, e compatível com as políticas e normas que regem a instituição e a sua atividade fim.



Desenvolver políticas de prevenção e procedimentos específicos de contenção de danos, avaliação de impacto e mitigação de riscos a fim de assegurar o bem-estar, a segurança e os direitos de todos os envolvidos.



Assegurar que existam medidas de segurança implementadas para proteger os dados dos usuários contra acessos não autorizados ou vazamentos, estando atento às legislações mais relevantes de proteção, como o GDPR na Europa e a LGPD no Brasil, garantindo privacidade e transparência dentro do ambiente da instituição.



Sempre que possível, evitar utilizar dados pessoais de terceiros. Em caso de estrita necessidade de utilizar dados de terceiros em sistemas que contenham modelos de IA, assegurar que as pessoas envolvidas tenham ciência previamente e tenham a opção de concordar ou não com a utilização de seus dados.

-  Promover a integração do uso de IA com as práticas de inovação, segurança e proteção de dados e conduta ética no âmbito da instituição.
-  Estimular a utilização responsável de IAG no ambiente – acadêmico, de pesquisa e administrativo – através de campanhas de conscientização sobre o uso de IAG por todos os profissionais e estudantes.
-  Manter-se informado o quanto possível sobre o avanço de marcos regulatórios e legislações específicas sobre usos e implicações da IAG na educação. Participar e incentivar a comunidade a participar de eventos e seminários que se debruçam sobre o tema. A estagnação no aprendizado pode levar ao uso inadequado ou obsoleto da tecnologia.
-  Garantir que as ferramentas de IA, se usadas, sejam acessíveis a todos os alunos, independentemente de diferenças sociais ou econômicas, adaptando a tecnologia para responder às necessidades de aprendizagem que consigam promover o justo equilíbrio de oportunidades nos ambientes educacionais.
-  Não evitar conversas difíceis sobre as implicações éticas da IAG. Ignorar esses diálogos pode resultar em uma compreensão limitada das responsabilidades éticas associadas ao seu uso. Promover discussões abertas conectando outros pesquisadores sobre os desafios e oportunidades. Isso inclui explorar, independente da área de pesquisa, questões de viés, privacidade e o impacto futuro da IA na sociedade.
-  Incorporar o uso de IAG como parte do sistema de governança da instituição, acompanhando os tipos de uso e, sempre que possível, resultados obtidos pelos modelos utilizados.
-  Ao criar grupos e comitês, garantir uma composição diversificada e inclusiva, representando a pluralidade de disciplinas, perspectivas e experiências.
-  Manter as operações do comitê transparentes, divulgando abertamente suas metas, processos de decisão, achados e recomendações à comunidade.
-  Monitorar os usos e as ferramentas de IA regularmente para identificar, substituir ou corrigir quaisquer vieses que possam levar a tratamentos injustos ou discriminatórios.
-  Criar meios para estimular o compartilhamento de resultados obtidos com modelos de IAG – desejáveis ou não, positivos ou não.
-  Procurar entender a natureza de resultados indesejados e suas causas. Intolerância a qualquer tipo de problema ou resultado indesejado pode desestimular o uso da tecnologia ou, principalmente, o compartilhamento de experiências com a instituição.
-  Promover um diálogo institucional aberto, incluindo fóruns, seminários e consultas públicas. Isso permite o intercâmbio de ideias, preocupações e expectativas sobre o uso da IAG na educação e na sociedade.
-  Não ignorar ou minimizar as preocupações e contribuições de pesquisadores ou instituições em função de aparente prestígio institucional. A falta de engajamento realmente democrático pode levar a decisões elitistas e resistências às iniciativas propostas.
-  Desenvolver programas de educação e capacitação para membros e não membros dos comitês, expandindo os horizontes de informação e pesquisa, visando aumentar a compreensão sobre a IAG e seus impactos éticos, técnicos e sociais.

-  Não assumir que todos os membros do grupo estejam no mesmo nível de conhecimento ou interesse pelo tema.
-  Fortalecer a interlocução entre os membros independentemente do nível de formação, uma vez que a ausência desse intercâmbio pode inibir o aperfeiçoamento educacional e comprometer o próprio propósito do comitê.
-  Buscar análises críticas (inclusive autocríticas) e reflexivas. Não adotar uma abordagem exclusivamente tecnocrática, que desconsidere as complexas implicações éticas e humanas sobre o uso das ferramentas.
-  Encorajar uma análise destemida que se ocupe dos reais impactos da IAG, considerando tanto os benefícios quanto os riscos. O grupo deve ser plural o suficiente para ponderar sobre questões éticas, legais e sociais, formulando recomendações que considerem o benefício da comunidade.
-  Manter o comitê em contato ativo com outros grupos, reforçando parcerias institucionais, ampliando o diálogo entre os vários interessados e ajustando suas estratégias, quando necessário. A tecnologia está em constante evolução e as abordagens precisam ser flexíveis para se adaptarem a novos momentos. Indicar claramente a utilização de IAG como parte do processo de produção e desenvolvimento de atividades.
-  Não se apegar, portanto, a conceitos ou políticas inflexíveis. Resistências puramente personalistas podem impedir o grupo de responder eficazmente aos desafios e oportunidades apresentados pela IAG.
-  Indicar claramente a utilização de IAG como parte do processo de produção e desenvolvimento de atividades.
-  Um modelo de IAG pode gerar resultados inesperados ou indesejáveis. Praticar a análise crítica antes de seguir adiante com qualquer resultado gerado pelos modelos de IAG.
-  Promover a independência intelectual dos alunos ao incentivá-los a usar a IA como uma ferramenta de apoio, não como substituta do esforço pessoal e do pensamento crítico, encorajando o trabalho analítico rigoroso que sempre considere os impactos do uso da IA na própria aprendizagem.



**EXTRA DOSSIÈS**

LE PENSEUR

# Criatividade da Inteligência Artificial Generativa

Ana Maria Di Grado Hessel<sup>1</sup>

David de Oliveira Lemes<sup>2</sup>

**Resumo:** O texto tem como objetivo refletir sobre a criatividade com o uso da IA generativa. A temática tem sido muito debatida em função das opiniões especulativas que ora valorizam o potencial criativo das ferramentas ou recursos e ora depreciam sua utilidade. O conceito de criatividade humana é apresentado na visão de alguns autores e pode ser sintetizado como a capacidade de produzir algo original ou inovador. O texto discute se a IA generativa pode simular a criatividade humana. Argumentos são colocados para mostrar que os processos de criação do homem e da IA generativa são diferentes, pois os algoritmos não são capazes de desenvolver algumas lógicas humanas. Entretanto, a IA generativa pode gerar produtos aceitáveis e interessantes na medida em que o prompt criado pela inteligência humana, para fomentar a geração de um produto, for bem construído. A semelhança entre os produtos da IA generativa e os trabalhos humanos destaca o potencial dessa tecnologia em expandir as fronteiras da criatividade.

**Palavras-chaves:** criatividade; Inteligência Artificial Generativa; prompt; produto; inovação

---

<sup>1</sup> Ana Maria Di Grado Hessel – Doutora e mestre em Educação/ Currículo pela PUCSP. Professora credenciada no Programa de estudos Pós-Graduados em tecnologias da Inteligência e Design Digital. Pesquisadora do GEPEC e GPTED. Orcid: <https://orcid.org/0000-0003-4776-7754>.

<sup>2</sup> David de Oliveira Lemes – Doutor em Tecnologias da Inteligência e Design Digital pela PUCSP. Diretor da Faculdade de Estudos Interdisciplinares da PUCSP, professor do curso de Ciência da Dados e Inteligência Artificial da PUCSP. Orcid: <https://orcid.org/0000-0003-0291-807X>.

## **Generative Artificial Intelligence creativity**

**Abstract:** The text aims to reflect on creativity using generative AI. The topic has been much debated due to speculative opinions that sometimes value the creative potential of tools or resources and sometimes depreciate their usefulness. The concept of human creativity is presented in the view of some authors and can be summarized as the ability to produce something original or innovative. The text discusses whether generative AI can simulate human creativity. Arguments are put forward to show that the processes of creation of man and generative AI are different, as algorithms are not capable of developing some human logic. However, generative AI can generate acceptable and interesting products as long as the prompt created by human intelligence, to encourage the generation of a product, is well constructed. The similarity between the products of generative AI and human works highlights the potential of this technology to expand the boundaries of creativity.

**Keywords:** creativity; Generative Artificial Intelligence; prompt; product; innovation

## Para começar a conversa

A IA generativa (IAG) é uma área da inteligência artificial que se dedica em criar soluções, conteúdos e dados novos, a partir de informações armazenadas em grandes bases de dados. Vale notar que a criação executada pela IA generativa tem como base os produtos que já foram concebidos pela mente humana. São vários os conteúdos gerados e incluem diversas áreas, como artes, músicas, textos, imagens e design de produtos. O processo de criação por IA generativa envolve várias técnicas e modelos, entre os quais os mais destacados são as redes neurais generativas e os modelos de linguagem como o GPT (Generative Pre-trained Transformer).

Como um modelo de linguagem, a IA generativa tem sido experimentada e tem sido implementada em várias plataformas e interfaces, permitindo aos usuários interagirem por meio de websites, aplicativos, e integrações em outros *softwares*. São muitas as apreciações contraditórias a respeito do seu uso, suscitando opiniões favoráveis e temores de toda a natureza. Entretanto, uma questão se mantém presente no âmbito das especulações: Em que medida a IA generativa pode contribuir para a criatividade humana?

A relação entre tecnologia e criatividade humana é um tema complexo e tem sido muito debatido. Enquanto alguns argumentam que a IAG pode estimular a expressão da criatividade ao oferecer novas ferramentas e recursos, outros alertam que ela pode gerar dependência e reduzir a inventividade.

A IA generativa está em constante evolução, com novos modelos e técnicas sendo desenvolvidos regularmente. Ela tem o potencial não apenas de automatizar tarefas criativas, mas também de abrir novas possibilidades para a criatividade humana, oferecendo ferramentas que podem inspirar e ampliar a expressão criativa. A favor da Inteligência Artificial Generativa, argumenta-se que ela pode gerar novas ideias em forma de texto e em forma de arte, além de proporcionar o aprendizado sobre diferentes culturas e suas perspectivas.

A aplicação da Inteligência Artificial Generativa abrange diversas áreas criativas, como o desenvolvimento de *softwares* de pintura, novos instrumentos musicais e formas de arte digital. Ela também pode ajudar na exploração de culturas e na criação de jogos educativos e simuladores de realidade virtual.

Entretanto, existe o risco da aplicação da IA resultar em produções artificiais e pouco originais, levando à confiança excessiva em tecnologia e à perda da capacidade de inovação. A criação de conteúdo enganoso também é uma ameaça e é crucial compreender os benefícios e riscos dessa tecnologia para seu uso responsável e ético. Alguns críticos ressaltam que a tecnologia poderia encorajar a inércia criativa e a confiança excessiva, criando conteúdo realista e persuasivo demais. A criação repetitiva e desprovida de inspiração também é uma preocupação.

Este ensaio tem como objetivo trazer elementos sobre a temática de modo a contribuir com a reflexão sobre a criatividade com o uso da IA generativa.

### **Como definir criatividade?**

O conceito de criatividade é multifacetado e pode ser definido de várias maneiras, mas geralmente envolve a capacidade de gerar ideias novas e originais. Isso pode incluir a produção de algo inédito e útil, o processo mental de geração de ideias, a combinação original de elementos existentes e a capacidade de transcender o convencional. A criatividade é vista como a habilidade de gerar ideias, soluções, ou produtos que são ao mesmo tempo originais e apropriados para uma determinada situação ou problema. A criatividade humana é caracterizada pela capacidade de estabelecer conexões inéditas entre ideias, conceitos ou objetos aparentemente não relacionados, resultando em algo novo, útil e valioso.

A criatividade não se limita à arte, campo na qual é valorizada, mas desempenha um papel importante em diversas áreas da atividade humana, sendo influenciada por fatores individuais, sociais, culturais e ambientais.

Um autor que discute criatividade é Mihaly Csikszentmihalyi (2013), um psicólogo húngaro-americano conhecido por seu trabalho sobre o conceito de “fluxo” e por suas contribuições significativas ao estudo da temática. Para o autor, a criatividade é definida como a produção de uma novidade. Não se trata da invenção de algo completamente novo, mas sim a capacidade de transformar e combinar ideias já existentes para produzir algo com originalidade.

Csikszentmihalyi (2013) destaca que a criatividade não ocorre no vácuo. Ela é reconhecida e valorizada dentro de um contexto cultural, social e histórico. Sua abordagem enfatiza o sistema de criatividade, que inclui a pessoa criativa e o conjunto de regras e práticas ou o domínio no qual

ela opera. Essa visão sistêmica da criatividade sublinha a importância das interações entre o indivíduo, a sociedade e a cultura na realização de contribuições criativas.

A criatividade desempenha um papel fundamental em diversas áreas da vida, incluindo educação, negócios, ciência, tecnologia, artes, entretenimento e solução de problemas sociais. Ela promove a inovação, a expressão individual, o pensamento crítico e a resolução de problemas de maneiras únicas e originais. Segundo Ostrower (1977), o potencial criador humano elabora-se nos múltiplos níveis do ser sensível-cultural-consciente do homem, e se faz presente nos múltiplos caminhos em que o homem procura captar e configurar as realidades da vida. Para o autor a criatividade se refaz sempre, de maneira inesgotável.

Fleith, Vilarinho-Resende e Alencar (2020) apresentam um modelo de criatividade concebido por Amabile. Essa psicóloga social, define criatividade como a produção de ideias ou produtos que são tanto originais quanto apropriados ao contexto no qual são desenvolvidos e enfatiza a importância da motivação intrínseca, do ambiente de trabalho favorável e do fluxo de trabalho para estimular a criatividade humana em todos os sentidos em busca da inovação.

A inovação pode ser definida como o processo de introdução de algo novo ou significativamente melhorado em termos de produtos, serviços, processos ou modelos de negócios, resultando em valor adicional para os indivíduos, organizações ou sociedade como um todo. Ela não se limita apenas à criação de algo totalmente novo, mas também envolve a aplicação criativa de ideias existentes de maneira única e eficaz. A inovação frequentemente surge da interseção entre a criatividade, que envolve a geração de ideias originais, e a implementação dessas ideias de forma prática e viável. E para Plonski (2017), “inovação é a criação de novas realidades”.

### **IAG pode simular a criatividade humana?**

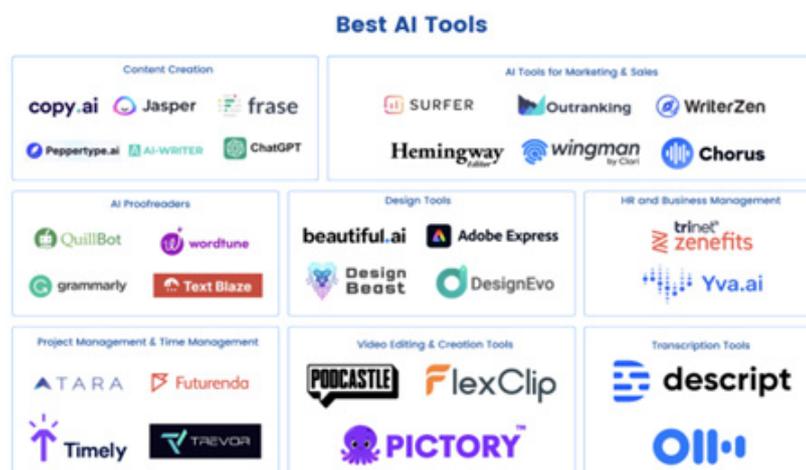
Do ponto de vista da IA, a criatividade é frequentemente descrita em termos de “criatividade computacional”. Isso se refere à capacidade de um sistema de IA de produzir resultados que seriam considerados criativos se originados por humanos. Isso inclui a geração de arte, música, textos literários, soluções inovadoras para problemas complexos, entre outros. No entanto, a criatividade em IA é limitada pelas informações, algoritmos e objetivos específicos programados por humanos. Pode-se observar se o sistema de IA é criativo se consegue gerar saídas que são novas, não vistas previamente ou incomuns dentro do contexto específico; se tem a capacidade de produzir saídas que são apropriadas, úteis e relevantes para

o problema ou contexto em questão; se tem a habilidade de modificar e combinar ideias existentes para criar algo distinto e original.

Ward (2020) explica que a criatividade computacional não precisa ser equivalente à criatividade humana, ou seja, as máquinas não precisam ser equiparadas aos humanos no quesito inventividade, mas podem contribuir com resultados considerados interessantes. Apesar dessas características, a criatividade da IA ainda é frequentemente vista como derivada das instruções e intenções humanas; as máquinas não possuem consciência ou motivações próprias. Assim, a criatividade de um sistema de IA é um reflexo tanto das suas capacidades técnicas quanto das visões e inputs humanos que o guiam. A criatividade computacional ainda enfrenta desafios, como a capacidade de entender e replicar ou reproduzir completamente os aspectos intuitivos, emocionais e contextuais da criatividade humana.

Santaella (2023) explica que as inteligências artificiais são distintas da inteligência humana. Questiona se o ChatGPT pode pensar, compreender e criar, frente às afirmativas de que o chat não é inteligente. Em sua narrativa a autora reúne elementos para explicar que os algoritmos não são capazes de desenvolver algumas lógicas humanas.

A IAG possui o potencial de ser um recurso valioso para a criatividade humana, oferecendo inspiração, aprimoramento e eficiência aos seus usuários. Ao gerar uma ampla variedade de ideias e estilos, a IA pode expandir os horizontes criativos dos indivíduos, encorajando a exploração de novas direções em seu trabalho. Além disso, ferramentas de IA podem ajudar os artistas a refinar suas ideias, aumentar sua produtividade e acessar recursos avançados que podem estar além de suas habilidades individuais, impulsionando assim a inovação e o progresso criativo. E, nos dias atuais, vemos uma grande variedade de ferramentas, acessíveis a um clique. Na fig 1 encontramos o logo de algumas delas.



**Figura 1.** Seleção de algumas ferramentas de inteligência artificial. Fonte: <https://ai-tools.tools/>.

Entretanto, apesar das oportunidades que a IA generativa oferece, também surgem preocupações sobre sua influência na originalidade e autonomia da expressão criativa humana. Questões éticas, como a autoria de obras geradas por algoritmos e o risco de estagnação criativa devido à dependência excessiva da tecnologia, devem ser cuidadosamente consideradas. Portanto, enquanto a IA generativa pode ser um ponto de apoio valioso para a criatividade humana, é essencial abordar essas preocupações de forma ética e colaborativa, garantindo um uso responsável e inclusivo da tecnologia.

O crescimento vertiginoso da aplicação de Inteligência Artificial Generativa aos processos criativos tem suscitado grandes debates acerca dos direitos autorais relacionados ao uso de sistemas dessa espécie. Assim, a utilização de Inteligência Artificial Generativa para a criação de obras literárias esbarra em questões complexas no que tange à autoria de obras produzidas por humanos. Isso pode gerar uma reestruturação do sistema de proteção de autores existentes no cenário brasileiro (Prass; Mügge; Bernasiuk, 2023).

### **Engenharia de prompt**

Um prompt é uma instrução ou entrada fornecida pelo usuário para direcionar a geração de conteúdo por parte do modelo ou serviço de IA. Essa instrução pode variar em complexidade e formato, dependendo da tarefa e do objetivo desejado. Um prompt pode ser tão simples quanto uma única frase ou palavra-chave, ou pode incluir informações mais detalhadas, como contexto, estilo desejado e restrições específicas.

Por exemplo, em uma tarefa de geração de texto, um prompt simples pode ser uma frase inicial que serve como ponto de partida para a IA criar uma história ou um parágrafo. Por outro lado, um prompt mais complexo pode incluir detalhes sobre o personagem principal, o cenário e o enredo desejado. Além disso, o prompt também pode conter instruções adicionais, como o tom ou estilo de escrita desejado, bem como quaisquer restrições ou requisitos específicos que devem ser seguidos pelo modelo de IA. Röhe e Santaella (2023) concluem no seu estudo que o comando mais elaborado pode fazer toda a diferença, mas os resultados alcançados nem sempre são atingidos de maneira confiável e satisfatória.

O guia “Recomendaciones para la docencia con inteligencias artificiales generativas” (2023), divulgado pela Universidad Carlos III de Madrid, sugere algumas formas de compor um prompt, com o objetivo de

obter um determinado resultado desejado: definir o papel de especialista que o chat vai assumir para compor a sua resposta; fornecer o contexto claro, com detalhes importantes para serem considerados na composição da resposta; explicar qual será público-alvo para o qual a resposta será dirigida; usar verbos claros e inequívocos; estabelecer claramente as restrições e limitações; definir o formato narrativo da resposta esperada. Depois de obtida a resposta, a sugestão complementar é iterar com o chat para refinar o texto, com o objetivo de alcançar o melhor patamar possível. Nesse sentido o prompt pode ser refeito para incorporar determinados detalhes.

A seguir apresentamos um exemplo de prompt usado para criar uma breve história, tendo como base uma ferramenta de inteligência artificial, o Copilot, desenvolvida para auxiliar usuários em várias tarefas, integrando-se com os produtos e serviços da Microsoft, encontrada no link <https://copilot.microsoft.com>

Prompt de entrada:

Crie uma história onde dois irmãos gêmeos, um menino e uma menina, de 9 anos de idade, estão perdidos em uma cidade grande como São Paulo (Brasil) e buscam um caminho para voltar para casa. Eles carregam um mapa da cidade, uma garrafa de água, uma mochila e ambos usam óculos e boné.

A história gerada:

Numa tarde ensolarada em São Paulo, os gêmeos Lucas e Luana encontraram-se numa situação inesperada. Após se distraírem com as vitrines coloridas de uma loja de brinquedos, eles se viram sozinhos em meio à multidão da grande cidade, longe dos olhares atentos de seus pais.

Com apenas 9 anos, mas munidos de coragem e um mapa da cidade, eles decidiram que era hora de encontrar o caminho de volta para casa. Lucas ajustou seus óculos e olhou para o mapa, enquanto Luana segurava firmemente a garrafa de água, sabendo que a aventura exigiria hidratação.

Eles caminharam por entre as ruas movimentadas, passando por vendedores de rua e músicos, cada um com sua mochila nas costas e um boné protegendo-os do sol. Em cada esquina, eles consultavam o mapa, comparando os nomes das ruas com os letreiros e buscando pontos de referência.

Quando a noite começou a cair, e as luzes da cidade se acenderam, os gêmeos se sentiram um pouco mais ansiosos. Mas então, eles avistaram algo familiar: a grande biblioteca onde sua mãe os levava nos fins de semana. Com os olhos brilhando de esperança, eles seguiram em direção à biblioteca, sabendo que dali para casa era um caminho conhecido.

Finalmente, após uma longa jornada, Lucas e Luana chegaram em casa, onde seus pais os aguardavam com abraços apertados e lágrimas de alívio. Eles haviam aprendido uma grande lição sobre responsabilidade e a importância de estar

sempre atento, mas também sobre a aventura e a independência que a grande cidade poderia oferecer. E assim termina a história dos gêmeos aventureiros, que com um mapa e coragem, descobriram que mesmo nas maiores cidades, há sempre um caminho para casa.

O mesmo prompt também foi usado para gerar imagens utilizando a mesma ferramenta, sendo que o comando inserido foi adaptado para gerar uma ilustração:

Crie uma ilustração onde dois irmãos gêmeos, um menino e uma menina, de 9 anos de idade, estão perdidos em uma cidade grande como São Paulo (Brasil) e buscam um caminho para voltar para casa. Eles carregam um mapa da cidade, uma garrafa de água, uma mochila e ambos usam óculos e boné. A ilustração deve ser feita no estilo cartum, preto e branco, desenhado com giz de cera.

A resposta da IA:

Acabei de criar a ilustração que você pediu: dois irmãos gêmeos, um menino e uma menina, de 9 anos, perdidos em uma cidade grande como São Paulo. Eles estão com um mapa da cidade, uma garrafa de água, uma mochila e ambos usam óculos e boné. A ilustração está no estilo cartum, preto e branco, e foi desenhada com giz de cera. Espero que goste! 😊

O resultado:



Figura 2: Imagens geradas pelo Copilot. Fonte: <https://copilot.microsoft.com>.

## O produto criado pela IA Generativa

A IA Generativa cria conteúdos por meio de modelos treinados em grande quantidade de conjuntos de dados. Esses modelos aprendem padrões, estilos, estruturas e relações dentro dos dados durante a fase de aprendizagem ou treinamento. Após treinados, eles podem gerar novos conteúdos que são similares em estrutura e estilo aos dados originais. O processo de criação por IA generativa envolve várias etapas e componentes críticos: na fase de treinamento o modelo de IA é exposto a uma grande quantidade de dados. Ao processar os dados, o modelo aprende padrões e relações. A partir de um prompt inicial, o modelo produz uma saída baseada no que aprendeu durante o treinamento, tentando fazer com que o conteúdo gerado seja coerente com o prompt fornecido.

O prompt, ou o comando utilizado para enviar uma solicitação em uma interface de IA generativa, tem um papel fundamental no processo de criação da resposta criada pela ferramenta. Ele serve como um ponto de partida para a geração, orientando o modelo sobre o que se espera como saída.

O teor do prompt é crucial, pois guia o processo criativo da IA influenciando diretamente a qualidade e relevância dos produtos gerados. A especificidade do prompt pode afetar significativamente o resultado produzido. Na medida em que o prompt é bem específico, os resultados obtidos são mais precisos e alinhados com as expectativas iniciais. O detalhamento com relação ao contexto, ao nível de informação e até mesmo a definição da área de conhecimento pode fazer a diferença na resposta ofertada pela ferramenta. Um prompt claro pode ajudar a evitar ambiguidades, orientando o modelo de forma mais efetiva para produzir o resultado desejado. Os resultados podem ter um teor original em função de instruções bem definidas.

No exemplo apresentado acima, gerado a partir de um prompt específico, podemos observar que a narrativa contém os elementos solicitados. Mas vai além, pois também avança, ambientando a situação com detalhes coerentes com o cenário e com o conto. Frente a veracidade do produto não é impossível supor que ele pode ter sido criado por um ser humano. Sabemos que os caminhos criativos da IA generativa e dos humanos são diferentes, mas os resultados produzidos podem se apresentar muito semelhantes.

## Conclusão

A IA generativa, com sua capacidade de aprender a partir de vastos conjuntos de dados e criar conteúdo novo que espelha os padrões, estilos e estruturas dos dados originais, representa um avanço significativo na tecnologia e na criatividade.

O papel central do prompt na orientação do processo criativo ressalta a importância da interação humana na definição dos resultados finais, evidenciando que, embora a máquina execute o trabalho de geração, a orientação humana é crucial para garantir a relevância e precisão do conteúdo produzido. Esse equilíbrio entre a capacidade computacional e o direcionamento humano abre novas possibilidades para a criação de conteúdo, permitindo resultados que, em muitos casos, são indistinguíveis dos criados diretamente por humanos.

A similaridade entre os produtos da IA generativa e os trabalhos humanos não apenas demonstra o avanço da tecnologia em simular processos criativos humanos, mas também destaca o potencial dessa tecnologia em expandir as fronteiras da criatividade, oferecendo ferramentas poderosas para explorar novos territórios criativos.

## Referências

- FLETH, Denise; VILARINHO-RESENDE, Daniela; ALENCAR, Eunice. O modelo componencial de criatividade de Teresa Amabile. In: NEVES-PEREIRA, Mônica Souza; FLEITH, Denise de Souza (orgs.). *Teorias da criatividade*. Campinas: Alínea, 2020.
- CSIKSZENTMIHALYI, Mihaly. *Creativity: Flow and the psychology of discovery and invention*. New York: Harper Collins, 2013.
- OSTROWER, Fayga. *Criatividade e processos de criação*. Rio de Janeiro: Vozes, 1977.
- PRASS, Rochele Moura; MÜGGE, Ernani; BERNASIUK, Helen Lentz Ribeiro. Autoria em tempos de inteligência artificial generativa: um olhar para a produção ficcional contemporânea no Brasil. *Revista Texto Digital*, v. 19, n. 2, p. 75-106, 2023.
- PLONSKI, Guilherme Ary. Inovação em transformação. *Estudos Avançados*, v. 31, p. 7-21, 2017. Disponível em: <https://www.scielo.br/j/ea/a/3Vmk8zqHbrVcgBwhMTyTC7d/?format=html>.
- RÖHE, Anderson; SANTAELLA, Lucia. IAs Generativas: a importância dos comandos para texto e imagem. *Aurora: Revista de Arte, Mídia e Política*, v. 16, n. 47, p. 76-94, 2023.

SANTAELLA, Lucia. *Há como deter a invasão do ChatGPT?* São Paulo: Estação das Letras e Cores, 2023.

UC3M. *Universidad Carlos III de Madrid*. Guia: Recomendaciones para la docencia con inteligencias artificiales generativas, 2023. Disponível em: <https://hdl.handle.net/10016/37989>. Acesso em mar 2023.

WARD, Megan. Victorian fictions of computational creativity. In: CAVE, Stephen; DIHAL, Kanta; DILLON, Sarah. (eds.) *AI Narratives: A history of imaginative thinking about intelligent machines*. Oxford: Oxford University Press, 2020, p. 144-164.

## A periferia é protagonista:

# o uso ético e transformador da inteligência artificial

Ana Maria dos Santos Rodrigues<sup>1</sup>

**Resumo:** Este artigo é mais do que um relato de estudo, é um ato. Um ato de romper o silêncio, de erguer a voz e ecoar as lutas interseccionais que, por vezes, são silenciadas e segregadas. É um alerta por mudança, um chamado à consciência e à visibilidade. Aqui, subverte-se a narrativa hegemônica da lógica colonial, que fragmenta e desvaloriza as diversas frentes de luta e busca construir pontes por meio da tecnologia, especialmente das Inteligência Artificial.

**Palavras-chave:** periferia; inteligência artificial colonial; hegemonia

---

<sup>1</sup> Mestranda no programa de Tecnologias da Inteligência e Design Digital TIDD na Pontifícia Universidade Católica de São Paulo PUC-SP. Pós graduanda em Gestão Escolar pela USP / Esalq, bolsista. Pedagoga formada pela Universidade Presbiteriana Mackenzie. Pesquisadora colaborativa em grupos de pesquisa atuando principalmente nos seguintes temas: Educação, Inteligência Artificial, Antirracismo Contato: [anarodrigues877@gmail.com](mailto:anarodrigues877@gmail.com).

## **Periphery is the protagonist: the ethical and transformative use of artificial intelligence**

**Abstract:** This article is more than a study report, it is an act. An act of breaking the silence, of raising one's voice to echo the intersectional struggles that are sometimes silenced and segregated. It is a cry for change, a call to consciousness and visibility. Here, the hegemonic narrative of colonial logic that fragments and devalues the various fronts of struggle is subverted, and bridges are built through technology, especially Artificial Intelligence.

**Keywords:** periphery; colonial artificial Intelligence; hegemony

## Introdução

Fazer a transição do silêncio à fala é, para o oprimido, o colonizado, o explorado, e para aqueles que se levantam e lutam lado a lado, um gesto de desafio que cura, que possibilita uma vida nova e um novo crescimento. Esse ato de fala, de ‘erguer a voz’, não é um mero gesto de palavras vazias: é uma expressão de nossa transição de objeto para sujeito-a voz liberta (Hooks, 2019, p. 38-39)

Baseado em fontes de pesquisa selecionadas, este artigo pretende apresentar uma visão crítica e propositiva sobre a epistemologia periférica e as tecnologias. D’Andrea (2020, n.p.) afirma que a “epistemologia periférica, como conceito, se constitui por meio de uma vivência que produz identificação com os sujeitos da pesquisa, oriundos da mesma classe social e com códigos compartilháveis. O cientista, quando lastreado por essa vivência compartilhada, há de realizar escolhas.” A vivência compartilhada entre o pesquisador e os sujeitos da pesquisa, ambos oriundos da mesma classe social, permite ao cientista compreender as escolhas e motivações dos participantes de forma mais profunda e autêntica.

Essa “voz liberta”, símbolo da emancipação de comunidades historicamente marginalizadas, encontra nas tecnologias, especialmente na Inteligência Artificial (IA), um terreno fértil para o florescimento de novas possibilidades. Buscamos, a partir dessa reflexão, destacar as possibilidades de mudança na promoção de iniciativas que gerem empoderamento e emancipação por meio da tecnologia, particularmente da IA a partir da seguinte questão: como as IAs podem ser reapropriadas para servir a comunidades historicamente marginalizadas?

O digital pode ser colocado a serviço da luta emancipatória? A IA pode ser reformatada e pensada para assegurar o interesse das classes populares, das comunidades tradicionais, ou está intrinsecamente vinculada à eficácia e à eficiência exigidas pelo capital? (Faustino; Lippold, 2023, p. 19).

A liberação da dualidade entre o bem e o mal, o atingível e o inatingível joga luz sobre o campo de batalha onde se desenrolam as lutas pela mudança, distante do hábito de enxergar essa dualidade nas IAs, pois, para Santaella (2023, p. 14), “a contradição entre o bem e o mal é indicadora do pouco entendimento que reina sobre a IA”. Isso não significa negar que as inovações tecnológicas emergem como uma espada de dois gumes neste cenário: por um lado, têm o poder de catalisar mudanças sociais positivas, democratizando o acesso à informação, melhorando a saúde e a educação, e impulsionando a sustentabilidade. Por outro lado,

a mesma tecnologia pode ser utilizada para fins de alienação e controle, ampliando desigualdades e fomentando a desconexão entre indivíduos.

A interrogação filosófica não é mais saber como tudo é pensável, nem como o mundo pode ser vivido, experimentado, atravessado pelo sujeito. O problema é saber agora quais são as condições impostas a um sujeito qualquer para que ele possa se introduzir, funcionar, servir de nó na rede que nos rodeia. (Foucault, 1971, p. 30)

Neste ponto de inflexão, torna-se crítica a escolha entre adotar a tecnologia como um motor de mudança social positiva ou permitir que ela se torne um instrumento de alienação e opressão. Para caminhar na direção do progresso, é essencial cultivar uma mentalidade crítica e questionadora, aberta à inovação e à mudança, mas sempre consciente das implicações éticas e sociais de nossas escolhas tecnológicas. Para Kaufman (2022, p. 83), “colocam[-se] para a sociedade novos desafios éticos e a premência de estabelecer arcabouços legais”.

Para além de deduções hipotéticas em que colocamos, de um lado, as IAs como detentoras de um poder absoluto, capazes de manipular e substituir as pessoas, e, de outro lado, as periferias como meros receptores de desigualdade, a proposta é inverter a lógica dos papéis. Isso significa escolher um lado: o lado do “bem”, do “atingível” e do humano como ser consciente, ético e empoderado. A intenção não é fecharmos os olhos para cenários de exclusão digital e de condições básicas de vida subtraídos nas margens da periferia, mas sim reconhecermos como valor a capacidade dos indivíduos de desmantelarem tais impedimentos e agirem rompendo estruturas. O otimismo, a esperança não devem ser interpretados como ingenuidade neste estudo, mas sim como um pensar complexo, que vê o todo e a auto capacidade dos indivíduos em escreverem suas próprias histórias.

## Metodologia

Este artigo adota uma metodologia flexível, guiada pelo princípio de estar “aberta às emergências, ao novo, ao possível, ao acontecimento”, conforme destacado por Santos (2005, p. 140). Essa abordagem é essencial para navegar a complexidade dos fenômenos sociais contemporâneos, permitindo que o processo investigativo seja dinâmico e adaptável, capaz de acolher novas realidades e *insights* à medida que elas emergem durante a pesquisa.

Em conformidade com Antunes (2016, p. 329), a metodologia empregada busca “identificar quais são as forças atuantes nesse campo, as partes que se desagregam, aspectos de exclusão e de inclusão, deslocamentos e apropriações”. Isso significa que a análise vai além da superfície, explorando as dinâmicas subterrâneas que moldam os fenômenos em estudo. Ao fazer isso, a pesquisa se compromete a revelar as complexas interações entre diferentes elementos e agentes, iluminando tanto as estruturas de poder que perpetuam a exclusão quanto os movimentos de resistência e reivindicação por inclusão e reconhecimento.

A partir de um olhar crítico, a intencionalidade de visibilizar inovações de projetos que, com a colaboração das IAs, mergulham nos contextos periféricos, valorizando os saberes e experiências silenciadas, é fundamental para construir uma práxis tecnológica que não perpetue a colonialidade, mas sim, a questione. Quando colocada nas mãos de quem compreende suas necessidades e potencialidades, a tecnologia deixa de ser uma imposição e se transforma em ferramenta de transformação social e mudança.

A discrepância apontada por Santos (2007, p. 20), entre teoria e prática social, ressalta a importância de uma metodologia que não somente reconheça essa divisão, mas que também se esforce para superá-la. A abordagem aqui adotada procura unir teoria e prática de maneira produtiva, entendendo que uma “teoria cega” falha em perceber as realidades sociais enquanto uma “prática cega” não consegue se valer dos insights e das estruturas conceituais oferecidas pela teoria. Portanto, a metodologia não apenas se alimenta da teoria para compreender a prática social, mas também permite que as observações práticas informem e refinem a teoria, criando um ciclo virtuoso de conhecimento e ação.

## O ato de mudar

Como é estranho que a natureza da vida seja a mudança, ainda que a natureza dos seres humanos seja resistir à mudança. E quão irônico é o fato de que os tempos difíceis que tememos poderem nos arruinar sejam exatamente aqueles que podem nos quebrar por dentro e nos ajudar a florescer naquilo que estávamos destinados a ser. (Lesser, 1953)<sup>1</sup>

---

<sup>1</sup> How strange that the nature of life is change, yet the nature of human beings is to resist change. And how ironic that the difficult times we fear might ruin us are the very ones that can break us open and help us blossom into who we were meant to be.

A natureza da vida é a mudança, seja ela individual, social, tecnológica, e a disposição para abraçá-la nos conduz a um terreno complexo de reflexão sobre o poder, a resistência, e (re)existência. A mudança, frequentemente percebida como disruptiva, exige que questionemos paradigmas que foram estabelecidos ao longo de séculos. Mas, por que é tão desafiador alterar o curso de nossas ações, questionar o *status quo* e adotar novas perspectivas?

Freire (1987) compreende que existir humanamente significa “pronunciar o mundo e assim, modificá-lo”. Nessa perspectiva enxergamos que, ao invés de um “sujeito” fixo e acabado, somos seres em constante devir, em permanente construção e reconstrução. As nossas identidades se formam e se transformam ao longo da vida, através da interação com o meio social e cultural, das relações que estabelecemos e das experiências que vivenciamos.

Para Freire (1994, p. 40), a disposição de abraçar a mudança está intrinsicamente ligada ao nosso existir e ao otimismo crítico: “isto é, a esperança que inexiste fora do embate” Esse otimismo crítico é fundamental para que possamos enfrentar as adversidades e resistir, ao mesmo tempo em que buscamos (re)existir em meio às transformações. O autor amplia essa visão ao afirmar que “a palavra assume o sentido de dizer o mundo e fazer o mundo. Ou seja, palavra verdadeira é práxis social comprometida com o processo de humanização, em que ação e reflexão estão dialeticamente constituídas” (Freire, 1987, p. 77).

Essa concepção nos leva a entender que a mudança não é apenas algo a ser enfrentado, mas também uma oportunidade para reafirmarmos nosso compromisso com a humanização e a construção de um mundo mais justo.

Ao nos engajarmos na práxis de transformação, nos deparamos com o desafio de desvelar o mundo da opressão. Freire (2005, p. 45) destaca a importância desse desvelamento e do comprometimento dos oprimidos em sua própria libertação: “O desvelamento do mundo da opressão e o comprometimento, pelos oprimidos, na práxis, com a sua transformação; realidade opressora transformada, esta pedagogia deixa de ser do oprimido e passa a ser pedagogia dos homens em processo permanente de libertação.”

Trata-se de um processo que não é apenas sobre a transformação do indivíduo, mas sobre a mudança coletiva que transcende a opressão, movendo-se em direção à libertação. “A palavra liberdade serve para expressar uma tensão muito importante, talvez a mais importante de todas.

Alguém quer sempre partir, e quando o lugar para onde quer ir não tem nome, quando é indeterminado e não se vê nas fronteiras, o chamamos de liberdade” (Larrosa, 2012, p. 297).

Entender a importância da mudança no comportamento individual e coletivo em relação a estruturas sociais arraigadas, como por exemplo o racismo e como são vistas as comunidades periféricas, requer uma análise profunda de como as identidades são formadas e transformadas dentro de uma sociedade. “Uma pessoa não nasce branca ou negra, mas torna-se a partir do momento em que seu corpo e sua mente são conectados a toda uma rede de sentidos compartilhados coletivamente, cuja existência antecede à formação de sua consciência e de seus efeitos” (Almeida, 2018, p. 53).

No processo de mudança, a complexidade da condição humana, como destacado por Morin (2013, p. 24), não atua como um obstáculo, mas como um facilitador da ação. Reconhecer a complexidade da existência humana e das interações sociais nos dá uma compreensão mais profunda dos verdadeiros riscos e oportunidades envolvidos em nossas lutas.

O mesmo podemos falar quanto à nossa maneira de conviver com as Inteligências. Do mesmo modo é imperativo reconhecer que as máquinas apenas espelham os preconceitos inerentes à sociedade que as cria. Sendo assim, é necessário mudarmos nosso olhar sobre a relação sociedade e tecnologia. O debate sobre vieses raciais, de perpetuação de desigualdades e preconceitos pelos algoritmos, que surge através do conceito de “racismo algorítmico”, abre caminho para algumas hipóteses sobre a relação do humano-máquina e a quem tange a responsabilidade já que os algoritmos, por mais automatizados que sejam, são fruto de padrões socialmente determinados.

Se os códigos são, mesmo em sua tendente automação, padrões socialmente determinados, a expressão “racismo algorítmico” não tenderia a escamotear a autoria do racismo, transferindo-a para os códigos enquanto oculta seus programadores, estes, sim, humanos formados e informados por dadas relações sociais de poder? (Faustino; Lippold, 2023, p. 150)

A atenção para o perigo de despersonalizar o racismo, atribuindo-o a entidades abstratas como algoritmos e códigos, deixa de reconhecer a responsabilidade dos programadores humanos. Ao fazer isso, corremos o risco de ignorar o contexto social e as relações de poder que moldam a criação de tecnologias, potencialmente perpetuando desigualdades existentes.

É necessário alertar para certa colonização da vida pelas máquinas e pelos algoritmos, mas a pergunta que as pessoas nem sempre se fazem é: quem domina quem? Se a máquina domina o humano, ainda que por meio de uma servidão voluntária, quem domina a máquina? Em outras palavras se algoritmos macabros colonizam nosso cotidiano para captar dados e induzir nosso comportamento e nossa subjetividade, com que razão o fazem? Será correto atribuir razão e, portanto, status de sujeito ao algoritmo quando ele próprio é programado por alguém com vistas a obtenção determinados resultados? (Faustino; Lippold, 2023 p. 40-41)

É possível que os indivíduos mudem? Que as Inteligências mudem, sejam elas artificiais ou humanas, quanto à tomada de decisão e quando ao agir em sociedade para o bem-estar comum a todos? A resposta parece ser um retumbante sim. “Observando e discutindo essas relações de poder interseccionais, temos uma oportunidade significativa para transformar a consciência entranhada na IA, uma vez que ela é, em parte, um produto de nossa própria criação coletiva” (Noble, 2021, p. 61). A capacidade de mudança é inerente tanto aos seres humanos quanto às IAs, embora os mecanismos de adaptação e evolução sejam distintos para cada um. Mas será que queremos mudar?

### O pensamento periférico

A partir da dialética obra de Devés-Valdés (2014), pensar na periferia envolve uma transformação profunda da maneira como compreendemos e interagimos com o mundo à nossa volta. O “pensamento periférico”, conforme descrito, é “aquele emergido no seio de comunidades impactadas pela presença, bem próxima do novo centro, que transforma a tradicional compreensão do mundo que possuíam” (*ibid.*, p. 38). Esse pensamento propõe uma reavaliação de nossas posições geográficas e culturais em relação ao centro. O pensamento da periferia desenvolve-se nos ecossistemas intelectuais das periferias e pode ou não estar estruturado sobre a base do dilema de ser ou não ser como o centro, ou seja, ser parte ou não do pensamento periférico “propriamente dito”.

É um pensamento que não apenas desafia a tradicional hierarquização espacial e cultural, mas também abre caminho para uma compreensão mais inclusiva e abrangente da realidade global. Nesse contexto, Devés-Valdés (2014) nos leva a refletir sobre a “planetária” como uma necessidade de “pensar planetariamente a totalidade das periferias, pensar sua condição de periferia para deixar de ser periferia” (*ibid.*, p. 24). Tal abordagem não se limita a reconhecer as disparidades, mas visa transformá-las, promovendo um equilíbrio mais justo e sustentável entre centro e periferia.

Por outro lado, a metodologia do bem-estar sugerida na obra de Devés aponta para um caminho prático e teórico de como as periferias podem reconfigurar sua posição no mundo. A partir da ideia da construção de uma metodologia de bem-estar, “deve supor-se que um estudo tão amplo das regiões periféricas de algum modo deve contribuir para que se pense melhor” (*ibid.*, p. 32), enfatizando a importância de uma abordagem holística e inclusiva no tratamento das questões periféricas.

Essa metodologia não apenas procura entender e melhorar as condições de vida nas periferias, mas também visa a uma reestruturação do pensamento, permitindo que “pensar a partir da periferia é pensar a relação de alteridade, a alteridade do centro, é pensar de outro modo e é também pensar para sair da condição periférica [...] para localizar-se em alguma centralidade que ofereça a possibilidade de abandonar essa condição de párias da história” (p. 72). Assim, pensar planeticamente na periferia é um convite à reinvenção e à busca de um espaço de dignidade e visibilidade no cenário global.

O reconhecimento dos aspectos positivos, no nível das sociabilidades, não elimina efeitos do que se sente enquanto carência local [...] porém a periferia não é a violência ou a carência que a ela se associam, estas são apenas algumas de suas faces, com as quais, aliás, se fazem muitas aprendizagens. São aprendizagens forjadas no contexto e que, às vezes, implicam processos de sobrevivência. (Freires; Pereira, 2023, n.p.)

Dentro deste contexto, entender a periferia requer uma abordagem que valorize suas contribuições únicas ao contexto social e cultural mais amplo. As aprendizagens forjadas na periferia, muitas vezes nascidas da necessidade de sobrevivência, têm o potencial de oferecer soluções inovadoras para problemas globais. Essas soluções podem se basear na criatividade, na solidariedade e numa profunda compreensão das dinâmicas locais, elementos que frequentemente são subestimados pelos modelos de desenvolvimento centrados no “centro”.

### O algoritmo da periferia

Periferias, vielas, cortiços  
Você deve tá pensando  
O que você tem a ver com isso?  
(Racionais MC's, 2002)

Ao citar “Negro Drama”, mergulhamos no universo da periferia, um espaço repleto de narrativas de resistência, luta e, sobretudo, de uma busca incessante por dignidade e reconhecimento. As palavras dos Racionais MC's não apenas pintam um quadro vívido da vida nas comunida-

des marginalizadas, mas também questionam o ouvinte sobre sua relação com essa realidade. Este é o cenário do que nomeamos, como “algoritmo da periferia” que ganha vida, não como um simples conjunto de códigos, mas como uma metáfora para as soluções criativas e inovadoras desenvolvidas nas favelas para enfrentar desafios diários.

Esses “algoritmos” não são codificados em linguagens de programação, mas sim nas vivências, nas culturas, nas estratégias de sobrevivência e nas inovações sociais desenvolvidas nas margens da sociedade. São “códigos” baseados na criatividade, na resiliência e na capacidade de adaptação das pessoas que vivem em condições muitas vezes marcadas por desafios socioeconômicos significativos.” O conceito de periferia não caberia em nenhuma definição externa senão nos próprios cidadãos periféricos, que, marginalizados pelas centralidades geográficas, afetivas, sociais e/ou estéticas se deslocam cotidianamente rumo ao centro de suas vontades e necessidades” (Brandão, 2016, n.p.).

Nascidas da necessidade e da escassez, as favelas evoluíram em complexos ecossistemas de vida, trabalho e cultura. Ainda que enfrentem desafios significativos, como a falta de acesso a serviços básicos e a infraestrutura inadequada, as comunidades periféricas também são locais de inovação e resiliência. As periferias são espaços onde “indivíduos versáteis são capazes de se adaptar, aprender, crescer constantemente, reposicionando-se em um mundo em rápida metamorfose” (Fava, 2018, p. 114). A vida nas periferias é marcada por uma constante reinvenção do espaço e das possibilidades, onde a adversidade é frequentemente transformada em oportunidade:

Queremos saber,  
O que vão fazer  
Com as novas invenções  
Queremos notícia mais séria  
Sobre a descoberta da antimatéria e suas implicações  
Na emancipação do homem  
Das grandes populações  
Homens pobres das cidades  
Das estepes dos sertões  
Queremos saber,  
Quando vamos ter Raio laser mais barato  
(Gilberto Gil, 1976)



**Figura 1.** A periferia em 0 e 1. Fonte: Imagem criada pela ChatGPT a partir de um prompt da autora, que captura a fusão entre a vida em uma comunidade de periferia e os algoritmos representados por 0 e 1.

A apropriação das tecnologias na periferia na visão de Nemer (2021) são “processos em que os oprimidos se apropriam de tecnologias cotidianas – artefatos, operações e espaços tecnológicos – e as utilizam para aliviar a opressão em suas vidas [...] são o modo como as pessoas exercem agência e conscientização e se apropriam de tecnologias para se mobilizarem em direção à qualidade de vida que desejam” (Nemer, 2021, p. 27) .

Nemer (2021) faz um paralelo entre a obra *Pedagogia do Oprimido* de Paulo Freire (1987) e em seu livro *Tecnologias do Oprimido*, que retrata as tecnologias como mundanas, pois “trata-se de como os moradores da favela traziam suas esperanças para se apropriarem de maneira criativa e crítica de tecnologias (artefatos, processos e espaços) e de suas jornadas para se libertarem” (Nemer, 2021, p. 243-244). Nesta perspectiva a periferia e as comunidades são espaços de intervenções tecnológicas, no passo a passo de iniciativas, projetos e movimentos focam na visibilidade e empoderamento na tentativa de reconstruir histórias.

No seio do desalento das favelas desenvolve-se uma vitalidade extraordinária. O importante não é contentar-se com uma sobrevivência biológica para constituir manadas dóceis e passivas... Trata-se de uma criação, da reconstrução de uma sociedade humana pelo desvio e recuperação dos objetos e das forças da modernidade a partir dos valores culturais e dos laços residuais das comunidades tradicionais. (Latouche, 1996, p. 117)

O quadro abaixo traz, em síntese, alguns projetos e iniciativas que utilizam as IA dentro das periferias.

**Quadro 1 IA na Periferia** (Fonte: elaborado pela autora com base em Rodrigues, Silva e Hessel 2024).

Projeto	Resumo	Instituição/local
Alunos indígenas e da periferia aprendem sobre robótica e inteligência artificial, em Manaus	Esta iniciativa visa democratizar o acesso à tecnologia e ao conhecimento científico, proporcionando a esses estudantes ferramentas e habilidades essenciais para o século XXI. Por meio de aulas práticas e teóricas, os alunos têm a oportunidade de aprender sobre programação, construção de robôs e os princípios da inteligência artificial, promovendo não apenas a inclusão digital, mas também inspirando-os a seguir carreiras na área de tecnologia e engenharia.	Escola Municipal Antônia Medeiros e Escola Municipal São Sebastião II  Secretaria Municipal de Educação  <b>Manaus</b>
CRI.A – IA Favela	O CRI.A é o primeiro influenciador digital de favela criado por inteligência artificial. Essa iniciativa visa promover a inclusão digital e o empoderamento tecnológico das comunidades de favelas, oferecendo conteúdo educativo sobre ferramentas gratuitas úteis para o cotidiano, geração de renda e criação artística. Através desta inovadora figura de influenciador, busca-se garantir que os moradores das favelas tenham acesso e possam aproveitar as oportunidades que a tecnologia e a inteligência artificial oferecem.	Felipe Fernandes Henrique Fernandes  <b>São Paulo</b>
CrI.Ativos da Favela	O projeto que criou um curso sobre IA, com os seguintes temas: Universo da I.A. como introdução à inteligência artificial (IA); conceitos básicos e usos na vida cotidiana; e IA aplicada ao audiovisual.	CUFA. Favela Filmes  <b>São Paulo</b>

MAISiCODE	O MaisiCode é um projeto social que visa transformar a vida de jovens de baixa renda através do ensino de tecnologia, com o objetivo de "Reprogramar a Quebrada". Oferecendo cursos de programação gratuitos	<b>São Paulo</b>
Pesquisador, Fábio Campos, usa Inteligência Artificial para estimular a consciência social de jovens da Favela da Rocinha	Projeto que integra tecnologia e educação social, essa iniciativa visa capacitar os participantes com o conhecimento e as ferramentas possíveis para abordar e resolver questões sociais em sua comunidade. Utilizando a IA para analisar e apresentar dados sobre desafios locais, como desigualdade, violência e acesso à educação, o projeto encorajou os jovens a desenvolverem soluções criativas e tecnológicas.	TLTL – Transformative Learning Technologies Lab. da Universidade de Columbia  <b>Rio de Janeiro</b>
Prototipando a Quebrada	O projeto "Prototipando a Quebrada" (PAQ) é uma iniciativa social que visa conectar jovens de comunidades periféricas com o conhecimento e oportunidades no mundo da tecnologia. Fomenta o uso da Inteligência Artificial, com a ideia de os jovens entenderem a importância do uso e em como utilizá-la para o desenvolvimento de projetos	<b>Florianópolis</b>

Dentre os projetos citados, dialogamos com o idealizador do CRI.A, Felipe Fernandes e a gestora da comunidade de aprendizagem do projeto Prototipando a Quebrada (PAQ), Yohana Galvani.

Os casos apresentados revelam a capacidade criativa e inovadora das comunidades pretas e periféricas de trazerem soluções práticas a sua população. Contudo, não se pode romantizar as dificuldades enfrentadas por esta importante parcela da sociedade. Desta forma, o termo “tecnologias de sobrevivências” apresenta uma complexa reflexão sobre a validação dos saberes periféricos, sem ignorar o fato de que tais soluções são provenientes de profundos problemas sociais que refletem historicamente o abandono do Estado e a centralização de poder. (Rodrigues; Luz; Nascimento, 2023)

Nesses diálogos foram abordados motivação, objetivos e os impactos das iniciativas. Ambos indicam uma transformação na maneira como a tecnologia é desenvolvida, enfatizando a importância de tornar o acesso à tecnologia mais abrangente e envolver diretamente as comunidades menos favorecidas na criação tecnológica e na utilização e desenvolvimento da IA. Os projetos são iniciativas concretas de inovação realizadas nas áreas periféricas.



**Figura 2.** Fonte: CRI.A IA Favela, 2024.

O CRI.A é o primeiro influencer negro criado por IA (Figura 2). O projeto envolve a criação de vídeos educativos para ensinar as pessoas das comunidades a utilizarem ferramentas de IA de forma acessível, além de explorar oportunidades de renda extra com essas habilidades.

Felipe Fernandes, idealizador do projeto, compartilhou sua experiência pessoal, desde sua origem na periferia de São Paulo até sua incursão na publicidade e tecnologia. Ele narrou como sua experiência profissional o levou a perceber as disparidades tecnológicas entre o Brasil e outros países, especialmente em relação ao uso da IA. Ele explicou como a ideia do projeto surgiu ao criar uma arte com IA e como isso o inspirou a promover a inclusão digital nas comunidades periféricas. Quanto à sua ampliação, surge a possibilidade de oferecer cursos de formação dentro da periferia. No entanto, para atingir esse patamar, é necessário investimento.

Precisamos alcançar sustentabilidade financeira para que o projeto se justifique, pois, além do seu valor social, ele também é um negócio. É preciso que seja reconhecido como tal, caso contrário, não alcançaríamos a visibilidade e o alcance que temos hoje. Atualmente, estamos presentes em nove portais, abrangendo até mesmo o entretenimento. Isso se deve às questões relacionadas ao empreendedorismo, tecnologia e inovação. (Felipe Fernandes)



**Figura 3.** Logo PAQ (ver texto). Fonte: Prototipando a quebrada, 2024.

O PAQ (Figura 3), que tem como idealizador Jefferson Lima, é um projeto cujo aspecto central encontra-se no seu empenho em promover a inclusão digital. Ao atuar dentro dos centros de inovação, busca-se uma aproximação física e conceitual entre a periferia e os espaços tradicionalmente vistos como reservados para a elite tecnológica. Eles contam com duas unidades Palhoça–Pedra Branca e Florianópolis. “Nós levamos a periferia para dentro dos centros de Inovação” (Yohanna, PAQ).

Até o momento não há nenhum programa ou formação direcionado diretamente para IA, porém o projeto fomenta e entende a importância do uso da IA para o desenvolvimento. Em sua narrativa, Yohanna, informa que se faz necessário no momento o alinhamento de Wokshops e rodas de conversas para incentivar a formação desses jovens para a utilização de ferramentas de IAs para expandir o repertório de uso “Os alunos já utilizam ferramentas de IA, texto, imagem, eles, por exemplo, já criam

prompts criativos e projetos com o uso do ChatGPT. A ideia é que eles entendam a importância das ferramentas, aprendam a utilizá-las para, em seguida, partir para o desenvolvimento” (Yohanna Galvani, PAQ).

O público-alvo são jovens de 16 a 24 anos de baixa renda. A atuação do PAQ ocorre dentro dos centros de inovação, com o foco em levar a periferia para dentro desses centros. O impacto do projeto é visto através de empregabilidade e renda dentro das comunidades.

## Considerações

Tim-Tim, um brinde pra mim  
Sou exemplo de vitórias, trajetos e glórias.  
(Racionais MC's)

É importante reconhecer que a periferia não é apenas um lugar, mas também um conjunto de saberes e experiências que podem contribuir significativamente para o desenvolvimento social e tecnológico. A periferia protagonista, entre metáforas, pesquisas e canções, vale dizermos, não “dorme”, ela é a primeira que levanta, que acorda, que resiste e se reconstrói como um espaço onde:

relações igualitárias, que se dão entre pessoas que possuem iguais direitos e deveres. Essas relações implicam que todos possam ter vez e voz, que todos sejam reconhecidos em sua singularidade, onde as diferenças sejam respeitadas. E mais: as relações comunitárias implicam, também, a existência de uma dimensão afetiva, implicam que as pessoas sejam amadas, estimadas e benquistas. (Guareschi, 1996, p. 97)

Nesse processo projetos e iniciativas resgatam o processo de empoderamento e representação às margens de quem só enxerga violência e opressão. Como impacto transformador, destacamos:

- **Inclusão Digital e Emancipação:** O acesso e o domínio sobre as novas tecnologias, especialmente a IA, abre portas para a inclusão digital das comunidades periféricas, oferecendo-lhes a possibilidade de superar barreiras históricas de acesso à informação e ao conhecimento.
- **Habilidades para o Futuro:** Além de desenvolver habilidades técnicas em robótica e IA, os participantes desses projetos são empoderados com ferramentas essenciais para o mercado de trabalho do futuro
- **Inspiração para uma Nova Geração:** A jornada inspiradora de superação e criatividade demonstra o potencial ilimitado de jovens de comunidades marginalizadas.
- **Soluções para Desafios Reais:** As iniciativas incentivam os participantes a imaginar e criar soluções tecnológicas que respondam às necessidades específicas de suas comunidades, promovendo um desenvolvimento social mais autêntico e eficaz.

Este estudo não apenas subverte a narrativa dominante que relega a periferia a um papel secundário, mas coloca-a no centro da discussão sobre o desenvolvimento e a aplicação ética da tecnologia. Através de exemplos concretos e reflexões teóricas, evidencia-se que a periferia não é apenas um espaço físico, mas um berço de inovação, resistência e, acima de tudo, de humanidade.

As sujeitas e os sujeitos periféricos aqui apresentados constituíram-se historicamente por uma série de circunstâncias. Na atualidade, em um mundo que prega cada vez mais a prosperidade, o individualismo e o empreendedorismo, sujeitas e sujeitos periféricos seguem mobilizados propondo equidade radical entre os seres humanos e baseados em um pressuposto de bem-estar geral. (D'Andrea, 2020, n.p.)

Além disso faz-se necessário um chamado à ação para que pesquisadores, desenvolvedores de tecnologia e formuladores de políticas considerem a periferia a partir de teorias e práticas que atribuam ferramentas “para a construção de um novo olhar sobre as periferias brasileiras” (Visões Periféricas, 2007, p. 5). Devemos olhar para a periferia como espaços de diálogos, na urgência e a necessidade de repensar o uso da IA sob uma perspectiva ética e consciente.

## Referências

ALMEIDA, Silvio Luiz de. *O que é racismo estrutural?* Belo Horizonte: Letramento, 2018.

ANTUNES, Camila. Sobre o “fazer periferia”: experiências, narrativas e reflexões a partir de um estudo etnográfico. *Revista Grifos*, n. 41, 2016.

BRANDÃO, Rebecca. Por uma estética do deslocamento. *Agulha: Calendário Cultural*, Rio de Janeiro, n. 5, jul. 2017.

CGC EDUCAÇÃO, 2024. *Pesquisador usa Inteligência Artificial para estimular a consciência social de jovens da Favela da Rocinha*. Disponível em: <https://cgceducacao.com.br/>. Acesso em: 10 mar., 2024

CRI.A. IA Favela. Instagram @iafavela. Disponível em: [instagram.com/iafavela/](https://www.instagram.com/iafavela/). Acesso em: 5 mar., 2024.

CRI.ATIVOS DA FAVELA, 2024. Instagram @cri.ativosdafavela. Disponível em: [instagram.com/cri.ativosdafavela/](https://www.instagram.com/cri.ativosdafavela/) Acesso em: 5 mar., 2024.

D'ANDREA, Tiaraju. Contribuições para a definição dos conceitos periferia e sujeitas e sujeitos periféricos. *Novos Estudos CEBRAP*, v. 39, p. 19-36, 2020.

- DEVÉS-VALDÉS, Eduardo. *Pensamiento periférico: una tesis interpretativa global*. Buenos Aires: Clacso. Idea-Usach, 2014.
- FAUSTINO, Deivison; LIPPOLD, Walter. *Colonialismo digital: por uma crítica hacker-fanoniana*. São Paulo: Boitempo, 2023.
- FAVA, Rui. *Trabalho, educação e inteligência artificial: a era do indivíduo versátil*. Porto Alegre: Penso, 2018.
- FOUCAULT, Michel. Entrevista concedida a Sergio P. Rouanet e J. G. Merquio. In: FOUCAULT, M. *et al. O homem e o discurso: a arqueologia de Michel Foucault*. Rio de Janeiro: Tempo Brasileiro, 1982, p. 17-42.
- FREIRE, Paulo. *Pedagogia do oprimido*. Rio de Janeiro: Paz e Terra, 1987.
- FREIRE, Paulo. *Comunicação e extensão*. Rio de Janeiro: Paz e Terra, 1992.
- FREIRE, Paulo. *Cartas a Cristina*, São Paulo: Paz e Terra, 1994.
- FREIRES, Thiago; PEREIRA, Fátima. Pensar o lugar das periferias no desenvolvimento de trajetórias educativas: apontamentos a partir de um estudo narrativo com jovens do Brasil e de Portugal. *Revista Brasileira de Educação*, Rio de Janeiro, v. 28, p. 1-25, 2023.
- GIL, Gilberto. Queremos saber. In: O viramundo (Ao vivo) (Vol 2). Album, 1976.
- GUARESCHI, Pedro. *Psicologia social comunitária: da solidariedade à autonomia*. Rio de Janeiro: Vozes, 1996.
- hooks, bell. *Ensinando a transgredir: a educação como prática da liberdade*. São Paulo: Martins Fontes, 2017.
- KAUFMAN, Dora. *Desmistificando a inteligência artificial*. Belo Horizonte: Autêntica, 2022.
- LARROSA, Jorge. Palavras desde o limbo. Notas para outra pesquisa na educação ou, talvez, para outra coisa que não a pesquisa na educação. *Revista Teias*, Rio de Janeiro, v. 13, n. 27, 2012.
- LATOCHE, Serge. *A ocidentalização do mundo: ensaio sobre a significação, o alcance e os limites da uniformização planetária*, 2ª ed., Petrópolis: Vozes, 1996.
- LESSER, Elizabeth. *Broken Open: How difficult times can help us grow*. New York: Random House, 1957.
- LOBO, Nuno. *Alunos indígenas e da periferia aprendem sobre robótica e inteligência artificial, em Manaus*. Radio Rio Mar, 3 dez., 2022. Disponível em: <https://radioriomarfm.com.br/estudantes-indigenas-e-de-escolas-perifericas-de-manaus-aprendem-sobre-robotica-e-inteligencia-artificial/>. Acesso em: 5 fev. 2024.

MAIS1CODE. Disponível em: <https://mais1code.com.br/>. Acesso em: 10 mar., 2024.

MORIN, Edgar; VIVERET, Patrick. *Como viver em tempo de crise?* Rio de Janeiro: Bertrand Brasil, 2013.

NEMER, David. *Tecnologia do oprimido: desigualdade e o mundano digital nas favelas do Brasil*. Vitória: Milfontes, 2021.

PROTIPANDO A QUEBRADA. Disponível em: [protipandoaquebrada.org](http://protipandoaquebrada.org). Acesso em: 5 mar., 2024.

RACIONAIS MC'S. Negro drama. In: *Racionais MC's*. São Paulo: Cosa Nostra, 2002.

RODRIGUES, Ana Maria dos Santos; LUZ, Solange Ferreira; NASCIMENTO, Thiago Gomes. Tecnologia antirracista, perspectivas identitárias e sociais da inovação. Trabalho apresentado no 2013 *International Conference on Innovation, Management and Technology* (ICIM 2023), Sao Paulo, PUC-SP, 2013.

RODRIGUES, Ana Maria dos Santos; SILVA, Luciana Cunha Lauria da; HESSEL, Ana Maria Di Grado. Tecnologia e emancipação: uma análise foucaultiana da Inteligência Artificial em contextos periféricos. *Revista Tecnologia Educacional*, Rio de Janeiro, n. 240, p. 19-32, 2024.

SANTAELLA, Lucia. *A inteligência artificial é inteligente?* São Paulo: Almedina, 2023.

SANTOS, Boaventura de Sousa. *Renovar a teoria crítica e reinventar a emancipação social*. São Paulo: Boitempo, 2007.

SANTOS, Edméa. *Educação online: cibercultura e pesquisa – formação na prática docente*. 2005. 351 f. Tese (Doutorado) – Curso de Educação, Faculdade de Educação, Universidade Federal da Bahia, Salvador, Bahia, 2005.

VISÕES PERIFÉRICAS. *Festival Audiovisual*. Rio de Janeiro, 2007.



**RESENHA**

LE PENSEUR

# Resenha de *Ética na inteligência artificial*, de Mark Coeckelbergh<sup>1</sup>

Dora Kaufman<sup>2</sup>

## A ética em pauta

Mark Coeckelbergh é um autor contumaz, com valiosas contribuições ao debate contemporâneo. No livro *Green Leviathan or the Poetics of Political Liberty: Navigating Freedom in the Age of Climate Change and Artificial Intelligence*, de 2021, Coeckelbergh recorre à ficção para refletir sobre as mudanças climáticas em um mundo mediado pela inteligência artificial (IA). Libertado da prisão após 20 anos, militante de um “futuro verde” constata encantado que seu sacrifício não foi em vão: a humanidade e a natureza foram salvas! A catástrofe climática, contudo, foi evitada com a IA controlando e manipulando o comportamento humano, ou seja, o custo foi a liberdade. Abordando diversos temas relacionados à crise climática, o autor deixa claro, contudo, que a solução não é a supressão da democracia.

No ano seguinte, 2022, Coeckelbergh publicou dois livros. Em *The Political Philosophy of AI* (2022, MIT Press), Coeckelbergh pondera que a IA coloca em xeque os significados de conceitos políticos-filosóficos tradicionais como liberdade, igualdade, democracia e poder, diante de “uma forma de capitalismo bruto, que expropria a

---

<sup>1</sup> COECKELBERGH, Mark. *Ética na inteligência artificial*. Tradução: Clariss de Souza *et al.* São Paulo / Rio de Janeiro: UB / Editora PUC-Rio, 2023.

<sup>2</sup> Dora Kaufman é professora do Programa Tecnologias da Inteligência e Design Digital da PUC SP. Doutora pela ECA-USP com estágio sanduíche na Université Paris Sorbonne IV, possui dois pós-doutorados: na COPPE-UFRJ e no TIDD-PUC SP. Autora de vários livros, entre eles *A inteligência artificial irá suplantará a inteligência humana?* e *Desmistificando a inteligência artificial*. Colunista da Época Negócios e colaboradora do Globo e do Valor Econômico. ORCID: <https://orcid.org/0000-0001-7060-4887>

experiência humana e impõe um novo tipo de controle: concentração de conhecimento significa concentração de poder”, com as grandes corporações moldando o futuro da sociedade.

O argumento central do livro *Self-Improvement: Technologies of the soul in the age of artificial intelligence* é que o autoaperfeiçoamento não é mais opcional, dado que os algoritmos de IA são parte essencial da cultura de autoaperfeiçoamento, além de um imperativo na medida em que as tecnologias digitais não nos oferecem apenas informações, mas nos convidam a nos comparar uns com os outros. Essa constante comparação “nos arrasta para regimes de autodisciplina, autovigilância e conhecimento quantitativo duros e insuportáveis”. Com a IA, nosso self é quantificado, tornando-se um “data-self”. Coeckelbergh advoga a favor de modelos que incentivem a percepção de que o crescimento individual sadio é relacional, transformando os “outros” não em competidores, mas em parceiros.

Poucos temas escapam à investigação e reflexão de Coeckelbergh. No artigo *Can machine create art?*, de setembro de 2017, por exemplo, ele oferece uma estrutura conceitual para o debate filosófico sobre o status da arte da máquina e da criatividade da máquina, defendendo que devem ser consideradas formas não humanas de criatividade. Seria o algoritmo de IA um agente artístico? Coeckelbergh nos faz refletir sobre a natureza da arte e da criatividade humanas.

Entretanto, a contribuição mais efetiva de Coeckelbergh talvez esteja no livro *Ética na inteligência artificial*, publicado em primeira edição na “Essential Knowledge Series” da MIT Press em 2020. O livro aborda temas desde o histórico ao cultural e técnico da IA, contendo um bom número de referências, mesmo que parte dessas referências não sejam devidamente comprovadas pelos autores citados, comprometendo os argumentos construídos a partir delas. Coeckelbergh, como professor de filosofia da mídia e da tecnologia na Universidade de Viena, tem legitimidade para versar sobre uma visão geral da ética da IA; além de professor, ele é membro de vários conselhos consultivos de ética para robótica e IA, incluindo conselhos consultivos políticos como o “High-level Expert Group on Artificial Intelligence” constituído pela Comissão Europeia no início do processo de regulamentação da IA em 2018.

O livro inicia com uma discussão sobre a possibilidade da IA geral, ou seja, sistemas de IA com cognição no nível humano, na qual o autor não assume uma posição explícita, mas apresenta diversas visões filosóficas com o pressuposto de que no fundo do debate sobre IA estão

divergências profundas sobre a natureza do ser humano, inteligência humana, mente, compreensão, consciência, criatividade, significado, conhecimento humano, e ciência, portanto com ligações a muitas disciplinas, incluindo matemática, engenharia, linguística, ciência cognitiva, ciência da computação, psicologia e até mesmo filosofia. Para ele, tanto o filósofo quanto o cientista de IA estão interessados em compreender a mente e fenômenos como inteligência, consciência, percepção, ação e criatividade. São abordadas as mudanças tecnológicas e seu impacto na vida das pessoas, além de transformações na sociedade e na economia, com base no entendimento de que a tecnologia é sempre social e humana, pois seu uso ocorre em um contexto social. Neste sentido a inteligência artificial não é apenas sobre tecnologia, mas igualmente sobre o que os humanos fazem com ela, como a usam, como a percebem e experimentam, e como a inserem em ambientes técnico-sociais mais amplos. Ou seja, o autor sugere pensar a ética, que diz respeito às decisões humanas, em uma perspectiva histórica e sociocultural.

Em seguida, Coeckelbergh aborda a questão do estatuto moral da IA, os temas do “agenciamento” e da “responsabilização” procurando descrever os argumentos que entrelaçam o debate sem tomar uma posição específica. Os sistemas de IA recebem o atributo de “agenciamento” no sentido de que executam ações no mundo, e essas ações têm consequências morais. Contudo, isso não significa necessariamente atribuir à IA o status de “agente moral”. O autor agrega diversas visões filosóficas sobre a atribuição de “agente moral” aos sistemas ou algoritmos de IA, desde questionamentos sobre a capacidade necessária para a agência moral, salientando que a IA é produzida e usada por humanos, portanto, a tomada de decisões morais em práticas tecnológicas é da competência humana.

Do outro lado do espectro, estão os que pensam que as máquinas podem ser agentes morais semelhantes aos seres humanos, inclusive afirmam que é possível e desejável dar às máquinas um tipo de moralidade humana, que podem até ser melhores do que os seres humanos no raciocínio moral pois são mais racionais e não influenciáveis por emoções (posição que pressupõe, do ponto de vista do autor, uma ideia equivocada a respeito da natureza da moralidade ao reduzi-la a seguir regras com o risco de gerar uma “IA psicopata”, perfeitamente racional e insensível às preocupações humanas porque carece de emoções). Por essas razões, pondera o autor, poderíamos rejeitar a própria ideia de agência moral plena, ou poderíamos tomar uma posição intermediária, a saber, temos que dar à IA algum tipo de moralidade, algo como “moralidade funcional”

proposta por Wendell Wallach e Colin Allen ou uma “moralidade estúpida”, não baseada nas propriedades humanas como propõem Luciano Floridi e J.W. Sanders.

Outra alternativa seria tornar a agência moral dependente de um nível mínimo de interatividade, autonomia e adaptabilidade, ou aplicar critérios não antropocêntricos para o agenciamento moral. De qualquer forma, a ênfase é em uma abordagem do status moral relacional e socialmente enraizada, que não é nem abstrata nem formalizante, nem é baseada em atitudes superiores e hegemônicas. A ética da IA, conseqüentemente, nos obriga a reconsiderar as nossas atitudes morais peculiarmente humanas e a questionar a própria natureza humana e o nosso futuro.

Para Coeckelbergh a atribuição de agente moral está diretamente associada à responsabilidade: se você tem efeito no mundo e nos outros, você é responsável pelas conseqüências. Recorrendo à “condição de controle” de Aristóteles, o autor argumenta que a IA pode realizar ações e tomar decisões que têm conseqüências éticas, mas não está ciente do que faz e não é capaz de pensamento moral e, portanto, não pode ser moralmente responsável pelo que faz. “As máquinas podem ser agentes, mas não agentes morais, uma vez que carecem de consciência, livre arbítrio, emoções, a capacidade de formar intenções e assim por diante”, logo o recomendado é preservar para os humanos a responsabilidade sobre as conseqüências.

Responsabilidade significa responsabilização e explicabilidade, se algo der errado precisamos de uma resposta e de uma explicação, contudo essas exigências não são factíveis com a natureza das redes neurais (“black box”, não explicabilidade, opacidade e outros termos que definem a não transparência de como as redes neurais chegam ao resultado). Diante da limitação técnica dos sistemas de IA baseados em redes neurais, Coeckelbergh pondera que “não se trata principalmente de explicar ‘como funciona’, mas de como eu, como ser humano de quem se espera que seja responsável e aja com responsabilidade, posso explicar minha decisão”, demanda cujo grau de sensibilidade é função do setor e da tarefa a ser compreendida pelo sistema de IA.

Em seguida, Coeckelbergh discute os impactos reais da IA atual (IA estreita) como privacidade e proteção de dados, manipulação, desinformação, totalitarismo, segurança e proteção, preconceito, futuro do trabalho e mudanças climáticas. Adicionalmente, versa sobre alguns aspectos do processo regulatório europeu da IA e, por fim, pondera sobre a abordagem antropocêntrica da IA.

Para o autor, a ética não deve ser vista como um tópico marginal que tem pouco a ver com sua prática tecnológica, mas como parte essencial dela, contemplando o acúmulo de riscos tecnológicos e o consequente crescimento das vulnerabilidades humanas, sociais, econômicas e ambientais. A ética precisa ser levada em consideração no estágio inicial do desenvolvimento da tecnologia, denominada de “ética by design”, superando o caráter vago e abstrato dos princípios éticos, enfrentando o desafio de construir uma ponte entre esses princípios éticos e legais abstratos generalistas e as práticas de desenvolvimento e uso da tecnologia em contextos específicos. Uma barreira a ser enfrentada é a falta de interdisciplinaridade e transdisciplinaridade, consequência da lacuna significativa na formação e no entendimento entre, por um lado, especialistas das ciências humanas e sociais e, por outro lado, especialistas das ciências naturais e da engenharia, tanto dentro quanto fora da academia: especialistas com formação em ciências humanas precisam se conscientizar da importância de pensar sobre as novas tecnologias, como a IA, e adquirir conhecimento básico dessas tecnologias e do que elas fazem. Por outro lado, cientistas e engenheiros precisam ser mais sensíveis aos aspectos éticos e sociais do desenvolvimento e uso da IA.

Com a premissa de que a tecnologia não é apenas um instrumento nem tampouco neutra, ou um recurso externo, mas molda nossas ações e nossas narrativas, Coeckelbergh defende que, se queremos um futuro para a IA diferente, precisamos de histórias diferentes e tecnologias diferentes. Se não gostamos de uma história específica sobre a IA, o que precisamos não é apenas rejeitar a história (muito menos ignorá-la), mas reescrever a história ou escrever uma nova história.



LE PENSEUR

## Diretrizes para autores – TECCOGS

A *TECCOGS – revista digital de tecnologias cognitivas* é um periódico do Programa de Pós-Graduação em Tecnologias da Inteligência e Design Digital (TIDD) da Pontifícia Universidade Católica de São Paulo (PUC-SP). As edições são semestrais e exclusivamente digitais, disponíveis em [pucsp.br/pos/tidd/teccogs](http://pucsp.br/pos/tidd/teccogs).

A **TECCOGS recebe artigos e resenhas de doutores ou de especialistas, mestrandos, mestres e doutorandos em coautoria com doutores.**

**Título, subtítulo, resumo** (com no mínimo 1000 e no máximo 2500 caracteres com espaços) e **palavras-chave** (de três a seis termos) do artigo deve aparecer em português ou espanhol (caso o artigo esteja escrito nessa língua) e, logo em seguida, traduzidos para o inglês.

O(s) **nome(s) do(s) autor(es)** deve(m) estar logo abaixo do subtítulo do artigo, acompanhado de uma nota de rodapé (escrita em fonte *Times New Roman* tamanho 11 pt, espaçamento simples) contendo currículo e biografia (formação, vínculo acadêmico, área de atuação e e-mail) com, no máximo, cinco linhas.

Cada artigo deve possuir no mínimo 20.000 e no máximo 50.000 caracteres com espaços.

Resenhas devem possuir no mínimo 8.000 e no máximo 13.000 caracteres com espaços.

O **corpo do texto** deve ser configurado em fonte *Times New Roman* tamanho 12 pt, espaçamento 1,5 linhas, parágrafo alinhado à esquerda, sem hifenização. **Citações diretas com quatro linhas ou menos** devem aparecer entre aspas (“”) incorporadas ao corpo do texto, indicando a fonte entre parênteses no modelo “(SOBRENOME [em maiúsculas], ano de publicação, p. [número da página])”, conforme a Norma Brasileira (NBR) 10520 (ago. 2002) da ABNT.

As **citações diretas com mais de quatro linhas** devem ter recuo à esquerda de 4 cm, sem aspas, com fonte *Times New Roman* tamanho 11 pt, espaçamento simples, parágrafo justificado e sem hifenização.

**Imagens** (fotografias, ilustrações, diagramas, tabelas, gráficos) precisam ter resolução de, no mínimo, 100 dpi/ppi (*pixels* por polegada) e devem estar integrados ao corpo do texto, com imagem e legenda centralizadas e fonte especificada (para imagens da *internet*: “Disponível em: “<site>”. Acesso em: “dia mês abreviado ano”).

O texto deve respeitar o **Novo Acordo Ortográfico da língua portuguesa**, vigente desde 2009. De acordo com a Base XIX da Nova Ortografia, termos como “Inteligência Artificial”, “Psicologia Cognitiva”, “Informática” e “Filosofia” (quando se trata da área de conhecimento) devem iniciar com maiúsculas. Segundo a política de direitos autorais da revista, os autores se responsabilizam pelos direitos de uso de todas as imagens.

Para elaboração de resumos, citações e referências, a revista segue as NBR 6023 (ago. 2002), 6028 (nov. 2003) e 10520 (ago. 2002) da ABNT. Não são permitidas notas de fim. Notas de rodapé devem ser usadas o mínimo possível, exclusivamente para adicionar observações pontuais, nunca para indicar referências bibliográficas. Em fontes da *internet*, a autoria do texto deve ser indicada entre parênteses, bem como o ano de publicação e endereço e data de acesso.

Todas as obras mencionadas nas referências devem estar citadas ao menos uma vez no texto e, do mesmo modo, toda e qualquer obra mencionada no texto deve constar nas referências.

Esta revista não aceita o **ChatGPT e outros simulares** como autores ou coautores. A autoria implica a responsabilidade sobre o conteúdo veiculado que se expressa na assinatura do autor. Quando o autor tiver empregado recursos da **Inteligência Artificial Generativa** para algum tipo de tarefa, é preciso indicar precisamente em uma nota ao final das referências.

A TECCOGS disponibiliza um arquivo formato .DOC que serve de *template* com instruções e exemplificações e estilos detalhados para escrever o artigo. [Baixe o modelo aqui](#).