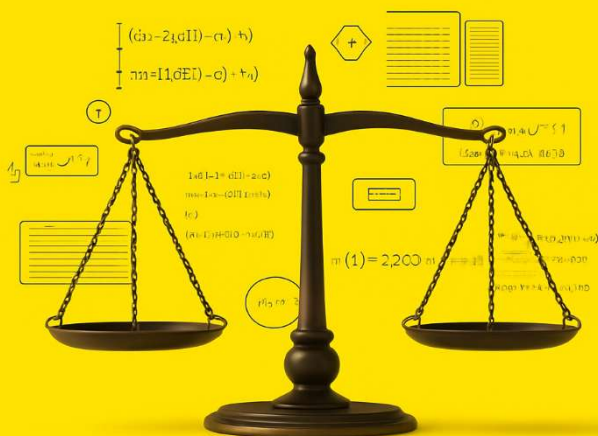


QUEM ESCREVE O CONHECIMENTO?



**Um Guia para o Uso Ético, Responsável e
Transparente da Inteligência Artificial Generativa
na Educação e na Pesquisa**

Paulo Roberto Córdova
Rosa Maria Vicari

**Paulo Roberto Córdova
Rosa Maria Vicari**

QUEM ESCREVE O CONHECIMENTO?

**Um Guia para o Uso Ético, Responsável e
Transparente da Inteligência Artificial Generativa
na Educação e na Pesquisa**



Copyright © 2026 dos Autores.

1ª Edição, janeiro de 2026.

Dados Internacionais de Catalogação (CIP)

Cordova, Paulo Roberto

Quem escreve o conhecimento? [livro eletrônico]: um guia para o uso ético, responsável e transparente da inteligência artificial generativa na educação e

na pesquisa / Paulo Roberto Cordova, Rosa Maria Vicari. -- 1. ed. -- Joinville, SC: Ed. dos Autores, 2026.

PDF

Bibliografia.

ISBN 978-65-01-88996-2

1. Educação - Finalidade e objetivos 2. Educação - Pesquisa 3. Inteligência artificial Aplicações educacionais 4. Inteligência artificial Aspectos morais e éticos I.
II. Título.

26-329480.0

CDD-371.334

Eliete Marques da Silva - Bibliotecária - CRB-8/9380

Índice para catálogo sistemático:

1. Inteligência Artificial : Educação 371.334.

Paulo Roberto Córdova.

Instituto Federal de Educação, Ciência e Tecnologia de Santa Catarina-IFSC.

E-mail: paulo.cordova@ifsc.edu.br

Rosa Maria Vicari.

Universidade Federal do Rio Grande do Sul-UFRGS.

E-mail: rosa@inf.ufrgs.br

Sumário

APRESENTAÇÃO	5
1. DESAFIOS PARA INTELIGÊNCIA ARTIFICIAL NA EDUCAÇÃO	8
1.1 Reflexões Iniciais	9
1.2 Produção ou Reprodução? Integridade Acadêmica Ameaçada	12
1.2.1 O problema da autoria.....	14
1.2.2 Propriedade intelectual (PI) e direitos autorais	18
1.2.3 O Problema da confiança: "alucinações" e vieses	20
1.2.4 Impactos da IAGen sobre a cognição e o aprendizado	22
1.3 Espelho da Sociedade: a quem a IA representa?	24
1.3.1 O mito do “humano genérico”	25
1.3.2 Implicações da homogeneização do pensar: a bolha do conhecimento da IA	27
1.4 Agentes de IA: uma nova fronteira	29
2. ÉTICA E GOVERNANÇA: necessidade urgente	32
2.1 Dimensões da Ética em IA.....	34
2.1.1 Ética em design (<i>ethics in design</i>)	34
2.1.2 Ética para o design (<i>ethics for design</i>).....	36
2.1.3 Ética por design (<i>ethics by design</i>)	37
2.2 Princípios para Ética em IA	38
3. USO ÉTICO DA IA NA PRODUÇÃO DE CONTEÚDO	45
3.1 A Erosão da Autoria: entre a produtividade e a responsabilidade intelectual.....	46
3.1.1 Riscos éticos centrais da IA generativa	48
3.2 Ferramentas Generativas como Apoio à Produção: limites e possibilidades.....	52
3.2.1 Por onde começar?	56
3.3 Conteúdos Audiovisuais: considerações específicas	62
4 TRANSPARÊNCIA E AUTORIA: um caminho a construir	64

4.1 Mandato de Transparência	64
4.2 Diretrizes Práticas para a Declaração de Uso de IA	65
4.2.1 AID framework	66
4.2.2 Simplificando as coisas.....	73
4.2.3 O que não fazer	79
GLOSSÁRIO	83
<i>DISCLOSURE</i> (DECLARAÇÃO DE USO DE IA)	85
REFERÊNCIAS	87
Sobre os Autores	93

APRESENTAÇÃO

A Inteligência Artificial consolidou-se como a força mais disruptiva na educação contemporânea. Não é um mero *upgrade* tecnológico, trata-se de uma redefinição fundamental de como o conhecimento é criado, transmitido e avaliado.

Das plataformas de aprendizado personalizado à emergência vertiginosa da inteligência artificial generativa — com ferramentas capazes de produzir textos, códigos e arte em escala —, a educação está sendo remodelada em seus alicerces.

Contudo, a potência dessa revolução tecnológica carrega riscos éticos proporcionais à sua capacidade de transformação. Este livro nasce da urgência de confrontar esses riscos. É imperativo que a comunidade acadêmica e educacional discuta de forma transparente e rigorosa como garantir a integridade, a equidade e a responsabilidade humana em um ecossistema cada vez mais mediado por algoritmos.

O conteúdo que se segue não propõe o banimento da IA. Pelo contrário: ele convoca à sua utilização consciente.

Neste livro, serão exploradas e discutidas questões fundamentais para que um avanço mais seguro e responsável possa ocorrer. Entre tais questões se destacam:

1. **A Crise da Autoria:** como a IA Generativa desafia a noção de trabalho intelectual original e qual é a responsabilidade do usuário (o "Mandato de Transparência").

2. **Vieses e Justiça:** de que forma os algoritmos, treinados em dados históricos enviesados, podem perpetuar ou ampliar desigualdades no ambiente educacional.
3. **Ameaça Cognitiva:** o risco real de dependência tecnológica e a perda de habilidades essenciais, como o pensamento crítico e o raciocínio analítico.
4. **O que Fazer Diante de Tais Transformações:** a partir de modelos, guias e recomendações internacionais, apresentamos orientações claras sobre quando e como declarar o uso de IA Generativa, incentivando práticas transparentes e eticamente fundamentadas na produção científica.

Nosso objetivo é delinear uma estrutura de Governança Ética que oriente o desenvolvimento e a aplicação da IA na educação e na pesquisa. Apresentamos os princípios fundamentais (como a transparência, a prestação de contas e a supervisão humana) e as diferentes dimensões da ética na IA, essenciais para que essa tecnologia seja uma forma de potencialização humana — e não uma substituta do nosso discernimento.

Este livro, portanto, é um guia e um manifesto. Ele reitera que a tecnologia deve estar a serviço dos mais altos valores da educação: a formação de cidadãos críticos e o desenvolvimento humano integral. Convidamos o leitor a participar ativamente deste debate. A escolha é definir, agora, as balizas éticas que garantirão que a IA construa um

futuro educacional mais justo e inteligente, sem nos tornar menos humanos.

1. DESAFIOS PARA INTELIGÊNCIA ARTIFICIAL NA EDUCAÇÃO

A Inteligência Artificial (IA) deixou de ser uma promessa futurista para se tornar uma força transformadora no presente, redesenhando setores, profissões e, fundamentalmente, a maneira como aprendemos, ensinamos e administramos a educação. Das salas de aula da educação básica aos laboratórios de pesquisa em Instituições de Ensino Superior (IES), os sistemas de IA oferecem um horizonte de possibilidades: personalização massiva do aprendizado, tutores virtuais disponíveis vinte e quatro horas por dia, automação de tarefas administrativas e analíticas preditivas para combater a evasão escolar e, mais recentemente, a geração de conteúdo sintético e resolução de problemas com a Inteligência Artificial Generativa (GenAI¹).

Contudo, essa revolução tecnológica não é neutra. Cada modelo de IA carrega consigo os valores, vieses e premissas de seus criadores. Com isso, sua aplicação no delicado ecossistema educacional exige um profundo senso de responsabilidade. A discussão ética sobre a Inteligência Artificial permeia todo o espaço educacional — da educação básica, ensino superior e formação profissional, à aprendizagem ao longo da vida.

Nesse contexto, escolas, redes de ensino, universidades, centros de formação e comunidades de aprendizagem enfrentam o desafio de integrar a IA de forma responsável e orientada por valores

¹ GenAI: a sigla adotada vem do nome em inglês *Generative Artificial Intelligence*.

humanos. Como espaços de construção do conhecimento e de desenvolvimento social, as instituições educacionais têm a dupla responsabilidade de, por um lado, explorar o potencial da IA para aprimorar processos de ensino e aprendizagem e, por outro, atuar como referências na adoção de práticas éticas, transparentes e inclusivas.

Com isso, a ética não deve ser vista como um obstáculo à inovação, mas como a bússola que assegura que a tecnologia permaneça a serviço do propósito maior da educação: promover o desenvolvimento humano integral, a equidade e a formação de cidadãos críticos e conscientes.

1.1 Reflexões Iniciais

A presença da IA na educação já é uma realidade multifacetada. Ferramentas de IA generativa como ChatGPT, Gemini, Qwen, Manus entre tantas outras, desafiam noções tradicionais de autoria e avaliação. Plataformas adaptativas ajustam o ritmo e o conteúdo do ensino às necessidades individuais de cada estudante. Sistemas de gestão acadêmica utilizam algoritmos para otimizar a alocação de recursos e prever tendências de matrícula.

Essa transformação digital impulsionada pela IA pode ser categorizada em três grandes áreas de impacto:

- **Apoio ao Ensino e Aprendizagem:** sistemas Tutores Inteligentes (STI), sistemas de recomendação de conteúdo,

plataformas de avaliação automatizada e ferramentas de feedback instantâneo são exemplos de aplicações de IA voltadas ao ensino e à aprendizagem. Esses recursos podem personalizar a aprendizagem, oferecer suporte adaptativo aos estudantes, otimizar o processo avaliativo e ampliar a capacidade de acompanhamento pedagógico por parte dos docentes. (Vicari *et al.*, 2023; Banco Mundial, 2024; Lemann, 2024; Unesco, 2024; Córdova, 2025).

- **Gestão e Administração Educacional:** No contexto da gestão e administração educacional, a IA tem sido aplicada em diferentes frentes, como na análise de dados para previsão de evasão estudantil, na otimização de horários e alocação de turmas, na automação de processos administrativos e de secretaria, e na gestão eficiente de recursos humanos e materiais. Essas aplicações visam aprimorar a tomada de decisão, reduzir custos operacionais e promover maior eficiência e transparência nas instituições de ensino (Banco Mundial, 2024; CIEB, 2024; NEES, 2024).
- **Acessibilidade e Inclusão:** Nesse âmbito, a inteligência artificial tem desempenhado um papel fundamental ao viabilizar ferramentas de tradução simultânea, sistemas de legendagem automática e tecnologias assistivas que removem barreiras comunicacionais, sensoriais e cognitivas. Esses recursos ampliam o acesso ao conhecimento, favorecem a participação de estudantes com deficiência e promovem

ambientes educacionais mais equitativos e acessíveis para todos. (OCDE, 2024; *U.S. Department of Education*, 2023; Vicari *et al.*, 2023).

Cada uma dessas aplicações, embora promissora, acende um alerta sobre a necessidade de diretrizes claras para sua implementação. A urgência deste debate reside na própria natureza dos sistemas de IA.

Até pouco tempo, predominava a ideia de que o uso da inteligência artificial na tomada de decisões representava um meio eficaz de eliminar preconceitos e vieses humanos, uma vez que as decisões seriam fundamentadas em dados supostamente objetivos e isentas de influências subjetivas como emoções, preconceitos e crenças. No entanto, com o tempo verificou-se que o processo de coleta, representação, seleção e utilização desses dados, assim como a forma de codificação dos algoritmos e de suas regras, pode facilmente incorporar — de diferentes maneiras — preconceitos de natureza pessoal, social e histórica nas tecnologias de IA (Casas-Roma *et al.*, 2021).

A questão central é que os algoritmos de inteligência artificial operam com base em dados, e os dados refletem o mundo como ele é, com todas as suas desigualdades e preconceitos. Sendo assim, modelos de IA treinado com dados históricos carregados de preconceitos discriminatórios podem perpetuar ou até amplificar

discriminações de gênero, raça, idade ou classe socioeconômica, com consequências devastadoras no contexto educacional.

Além disso, a proliferação de modelos de GenAI também introduziu uma mudança de paradigma no cenário educacional. Se a IA tradicional operava em tarefas auxiliares como análises e previsões, a GenAI possibilita a criação de conteúdo em escala e velocidade sem precedentes. Essa capacidade de simular a produção humana de texto, código e arte, levanta os dilemas éticos mais agudos e imediatos para o ecossistema educacional. O capítulo a seguir aborda essas questões com mais profundidade.

1.2 Produção ou Reprodução? Integridade Acadêmica Ameaçada

A popularização das inteligências artificiais generativas, a partir de 2022, marcou um ponto de inflexão no campo da tecnologia e do conhecimento. Essas ferramentas passaram a exercer influência significativa em múltiplas áreas, demonstrando capacidade de produzir textos com níveis de coerência, coesão e estilo comparáveis aos elaborados por seres humanos.

Ferramentas baseadas em GenAI, destacaram-se também, pela habilidade de sustentar diálogos extensos e contextualmente consistentes. Sua aplicabilidade rapidamente se expandiu, abrangendo atividades como síntese e revisão textual, classificação de textos, tradução automática, criação de resumos, preenchimento de

formulários e até o desenvolvimento e depuração de códigos de programação.

Todas essas tarefas podem ser executadas por meio de instruções em linguagem natural — chamadas de *prompts* —, o que torna o uso das inteligências artificiais generativas acessível a um público amplo, inclusive àqueles sem formação técnica. Essa característica democratiza o acesso a ferramentas de alta complexidade, permitindo que qualquer usuário realize atividades antes restritas a especialistas, ampliando, assim, as possibilidades de criação e automação no campo educacional e profissional.

Contudo, a popularização da GenAI também trouxe desafios significativos, especialmente no que diz respeito à integridade acadêmica e ao processo de aprendizagem. A facilidade com que um estudante pode gerar ensaios, resumos ou resolver problemas complexos em questão de segundos suscita reflexões éticas importantes sobre o uso indiscriminado e irresponsável dessas tecnologias. Tais práticas podem comprometer o desenvolvimento do pensamento crítico e da autoria intelectual, elementos fundamentais para uma educação ética e de qualidade.

Quando utilizada de forma acrítica, a GenAI pode levar à superficialização do aprendizado, substituindo o esforço intelectual pelo simples consumo de respostas prontas. Isso enfraquece habilidades essenciais como a análise crítica, a argumentação e a autonomia cognitiva — pilares fundamentais do processo educativo. Além disso, a dependência excessiva dessas ferramentas tende a criar

uma ilusão de competência, em que o estudante acredita dominar um conteúdo sem realmente compreendê-lo (Ahmad *et al.*, 2023).

No campo da integridade acadêmica, os riscos são igualmente preocupantes. A produção de textos gerados por IA sem a devida atribuição ou reflexão ética levanta um importante debate sobre a questão do plágio.

Por um lado, o uso acrítico da IA generativa pode levantar questões éticas, já que os modelos podem produzir textos inspirados em materiais protegidos por direitos autorais utilizados em seu treinamento, sem a devida atribuição. Por outro lado, a própria natureza da GenAI — que cria textos originais com base em incontáveis combinações e cuja origem das ideias é difícil de rastrear — torna complexa a acusação direta de plágio, diferentemente da cópia literal de fontes disponíveis em buscadores.

Diante dessa ambiguidade, as instituições de ensino são desafiadas a repensar seus métodos de avaliação, incentivando uma cultura de honestidade intelectual e promovendo o uso da IA como ferramenta de apoio ao raciocínio, e não como substituta do pensamento humano.

1.2.1 O problema da autoria

Em um cenário em que as máquinas são capazes de produzir resultados praticamente indistinguíveis do trabalho humano, torna-se indispensável refletir sobre o papel ético desempenhado por

estudantes e docentes no uso dessas tecnologias. Nesse contexto, destacam-se discussões que envolvem a autoria, a responsabilidade intelectual e a integridade acadêmica, temas que serão aprofundados ao longo deste capítulo por meio de uma análise crítica sobre os impactos e desafios da inteligência artificial generativa na educação.

1.2.1.1 Autoria compartilhada ou delegada?

Quando um texto é majoritariamente produzido por uma IA — por exemplo, 90% gerado pela máquina e apenas 10% editado por um humano — surge uma questão complexa: onde realmente reside a autoria? A noção tradicional de autoria, vinculada ao esforço intelectual e à originalidade humana, é colocada em xeque diante dessa nova dinâmica de produção textual.

Esse cenário exige que as instituições educacionais e científicas estabeleçam diretrizes claras sobre o uso aceitável da GenAI. É necessário diferenciar o emprego legítimo da IA como ferramenta de apoio — usada para estruturar ideias, aprimorar a escrita ou revisar textos — da delegação completa da tarefa, que transfere a responsabilidade criativa à máquina.

A ausência de critérios bem definidos pode gerar uma zona cinzenta entre autoria e coautoria, comprometendo tanto a transparência quanto a credibilidade das produções acadêmicas. Além disso, há implicações éticas importantes: quem responde por possíveis erros factuais, vieses ou violações de direitos autorais presentes em um texto gerado por IA?

Portanto, discutir autoria na era da GenAI não é apenas uma questão técnica ou burocrática, mas um debate sobre a própria natureza da criação intelectual. Reconhecer os limites da contribuição humana e os papéis das tecnologias inteligentes é essencial para preservar a integridade do conhecimento e valorizar o pensamento crítico em tempos de automação criativa.

1.2.1.2 A mentira da detecção

Depender exclusivamente de softwares capazes de “detectar” textos gerados por IA é uma estratégia frágil e, na maioria dos casos, ilusória. Esses sistemas são inerentemente falíveis: operam com margens de erro significativas, podendo tanto acusar injustamente produções humanas, quanto deixar passar conteúdos integralmente artificiais (Weber-Wulff *et al.*, 2023; Dalalah; Dalalah, 2023; Elkhataat *et al.*, 2023). Além disso, entram em uma espécie de “corrida armamentista” com a própria evolução dos modelos generativos, que se tornam cada vez mais sofisticados e difíceis de distinguir do texto humano (Ardito, 2023).

Essa dinâmica cria uma falsa sensação de controle e desvia o foco do verdadeiro problema: não é possível resolver uma questão ética apenas com ferramentas de vigilância tecnológica. Em vez de se limitar à repressão ou à tentativa de detectar o “uso indevido” da IA, a governança ética deve se concentrar na reformulação dos processos de avaliação, promovendo práticas pedagógicas que valorizem o

pensamento crítico, a autoria reflexiva e o uso responsável das tecnologias (Ardito, 2023).

Ao deslocar o eixo do controle para a formação, a escola e a universidade passam a tratar a IA não como uma ameaça a ser caçada, mas como uma realidade a ser compreendida, integrada e orientada eticamente dentro do processo educativo.

Em síntese, enfrentar os desafios técnicos, éticos e pedagógicos trazidos pela GenAI requer mais do que vigilância ou mecanismos de detecção: exige uma transformação no modo de ensinar. Educar não significa proibir, mas formar para a liberdade com responsabilidade. Cabe à escola e à universidade orientar o uso da IA de forma consciente, crítica e ética, ajudando os estudantes a reconhecerem tanto seus potenciais quanto seus limites.

Assim como a calculadora automatizou parte das operações mentais envolvidas nos cálculos — reduzindo o uso da memória de trabalho, da atenção e da visualização numérica — (Horodetska, 2025), a Inteligência Artificial começa a intervir em dimensões ainda mais complexas do raciocínio humano. Esse deslocamento cognitivo, mais profundo e abrangente, traz implicações formativas que a educação não pode ignorar (Kosmyrna *et al.*, 2025; Ahmad *et al.*, 2023).

Diante disso, é essencial buscar mais do que simples supervisão e monitoramento dos estudantes. É necessário promover práticas que fortaleçam o pensamento crítico, a autoria reflexiva e a responsabilidade ética, de modo que a IA amplie, e não substitua, a capacidade humana de compreender, analisar e criar.

1.2.2 Propriedade intelectual (PI) e direitos autorais

Um segundo pilar ético central da GenAI diz respeito à propriedade intelectual. Os modelos de IA são treinados em vastos *datasets* de dados disponíveis publicamente, incluindo obras protegidas por direitos autorais. Isso traz, ao menos, duas preocupações a serem consideradas, conforme discutido a seguir.

1.2.2.1 Violação de propriedade intelectual

O treinamento de modelos de Inteligência Artificial com obras protegidas por direitos autorais, sem o consentimento explícito de seus criadores, tornou-se um dos pontos mais controversos e judicializados no debate global sobre IA (NEES, 2024). Milhares de textos, imagens, músicas e outros conteúdos são incorporados a grandes bases de dados sem autorização, levantando sérias questões sobre violação de propriedade intelectual e exploração indevida do trabalho criativo humano (Unesco, 2023).

No contexto educacional, essa discussão assume uma relevância particular. Instituições de ensino que adotam tecnologias de GenAI precisam avaliar criticamente a origem dos modelos que utilizam — questionando se foram treinados de forma ética, transparente e em conformidade com a legislação de direitos autorais. Ignorar essa análise pode significar, ainda que indiretamente, legitimar práticas que ferem princípios éticos e jurídicos fundamentais (Australian Department of Education, 2023).

Mais do que uma obrigação legal, essa reflexão representa um compromisso com a integridade acadêmica e com o respeito à autoria.

Ao escolher conscientemente as ferramentas que empregam, escolas e universidades contribuem para um ecossistema educacional mais justo, em que a inovação tecnológica não se sobrepõe aos direitos e à dignidade de quem cria.

1.2.2.2 Entre a criação e o plágio

Essa apropriação maciça de conteúdos com direitos autorais para treinamento de IAs generativas nos leva também a uma profunda reflexão sobre a própria natureza da violação de propriedade intelectual no cenário digital, demandando novas categorias conceituais para descrevê-la. Nesse contexto, a violação de propriedade intelectual pode ser compreendida como uma forma de plágio em escala algorítmica.

Diferentemente do plágio tradicional, que se manifesta na cópia direta de um texto ou ideia sem o devido crédito, o plágio algorítmico ocorre quando uma IA reproduz ou gera conteúdos que são intrinsecamente derivados de obras humanas preexistentes, muitas vezes sem que seus autores originais sejam reconhecidos ou remunerados pelo uso de sua base criativa.

A fronteira entre criação e reprodução torna-se especialmente nebulosa e desafiadora nesse cenário. A criação humana autêntica pressupõe intencionalidade, reflexão crítica e originalidade — um processo subjetivo em que o autor transforma ativamente referências e experiências em algo genuinamente novo.

Em contraste, os modelos generativos de IA operam essencialmente por meio da recombinação estatística de padrões complexos, aprendidos a partir de imensos e massivos conjuntos de dados produzidos por outras pessoas. Assim, o que é apresentado como uma "criação inédita" pela máquina é, na realidade, uma síntese probabilística e sofisticada de criações anteriores, atuando como um mero eco da produção humana original.

Portanto, a busca por uma verdadeira inovação com o uso de inteligência artificial deve estar rigorosamente ancorada, não apenas na sua eficiência técnica ou capacidade produtiva, mas fundamentalmente no respeito inegociável à autoria e à ética da criação. Ignorar essa dimensão crítica e aceitar modelos treinados de forma passiva pode significar, ainda que de modo indireto, a legitimação de práticas que ferem princípios éticos, jurídicos e o valor intrínseco do trabalho criativo humano.

1.2.3 O Problema da confiança: "alucinações" e vieses

Conforme mostrado até aqui, a GenAI é capaz de produzir textos que soam convincentes, coerentes e gramaticalmente impecáveis. No entanto, é preciso deixar claro também, que essas produções podem veicular informações factualmente incorretas ou inteiramente fabricadas — um fenômeno conhecido como "alucinação". Conteúdos produzidos dessa forma são particularmente perigosos porque apresentam aparência de autoridade e verossimilhança, o que,

consequentemente, compromete o discernimento do leitor e o leva a validar um conteúdo destituído de sustentação empírica.

No contexto educacional e acadêmico, onde a precisão e a confiabilidade da informação são fundamentais, esse tipo de erro representa um risco significativo. Textos com dados falsos, referências inexistentes ou interpretações distorcidas podem comprometer o aprendizado, difundir desinformação e enfraquecer a credibilidade de produções científicas.

Nesse cenário, os problemas de maior recorrência a serem considerados incluem:

- **A fabricação de referências bibliográficas:** que consiste na criação de citações, títulos de obras e nomes de autores inteiramente fictícios. Este erro compromete a integridade acadêmica, pois induz estudantes e pesquisadores à citação de fontes inexistentes ou inválidas, fragilizando e comprometendo a fundamentação teórica e empírica de seus trabalhos.
- **A superficialidade conceitual:** que se caracteriza pela simplificação excessiva de conceitos teóricos complexos. Embora a fluidez textual seja mantida, a redução e simplificação do conteúdo prejudica o desenvolvimento de uma compreensão aprofundada e crítica dos temas, essenciais para o rigor científico e a formação intelectual.
- **A mistura de fatos verdadeiros e falsos:** que se refere à intercalação sutil de informações corretas com dados

factualmente incorretos. Esta mescla cria uma "zona cinzenta" de ambiguidade que se torna particularmente difícil de ser discernida e validada, inclusive por leitores com expertise na área.

- **A reprodução e amplificação de vieses socioculturais:** dada a natureza histórica dos *datasets* de treinamento, os modelos de IA demonstram a capacidade de replicar e exacerbar preconceitos latentes (por exemplo, vieses raciais, de gênero, sociais etc). Um exemplo notório é a associação estereotipada de determinadas profissões a um gênero específico em geradores de imagens. Impõe-se, portanto, às instituições educacionais a responsabilidade ética de mediar o uso dessas ferramentas e desenvolver a literacia crítica para a detecção e mitigação de tais vieses.

Portanto, o uso pedagógico da GenAI exige uma postura crítica e vigilante. Professores e estudantes precisam entender que a fluência verbal da IA não equivale à veracidade do conteúdo. Formar usuários capazes de validar, comparar e contextualizar as informações produzidas por essas ferramentas é um passo essencial para evitar que a educação se transforme em um repositório de erros bem escritos.

1.2.4 Impactos da IAGen sobre a cognição e o aprendizado

A crescente integração de tecnologias digitais nas atividades de estudo e pesquisa tem revelado um profundo impacto na forma como

os humanos aprendem, pensam e tomam decisões. Desde o uso de recursos assistivos comuns como calculadoras, até a aplicação de tecnologias inteligentes como motores de busca e aplicativos com GPS, o que inicialmente é tratado como conveniência, com o tempo acaba se tornando dependência (Dahmani; Bohbot, 2020; Musa; Ishak, 2021; Gong; Yang, 2024; Horodetska, 2025).

Para entender a complexidade desse fenômeno, é fundamental analisar as descobertas de dois estudos importantes que, embora distintos em suas abordagens, convergem em conclusões alarmantes.

O primeiro, conduzido com estudantes universitários, revela que o uso de ferramentas de GenAI tem contribuído para a perda da capacidade reflexiva e de tomada de decisões. A automação de tarefas intelectuais cotidianas, como a análise de dados e a elaboração de trabalhos, faz com que os indivíduos percam o interesse em realizar atividades por conta própria. Essa dependência crescente pode criar um comportamento de vício e levar à perda de habilidades como o raciocínio analítico e o pensamento crítico (Ahmad *et al.*, 2023).

Essa perda de habilidades e a dependência da IA se conectam diretamente com o conceito de "débito cognitivo" abordado em outro importante estudo realizado pelo Instituto de Tecnologia de Massachusetts (MIT). Esse estudo revela que, ao delegar tarefas intelectuais à IA, os humanos podem reduzir a atividade cerebral em regiões essenciais para a atenção, a memória e o raciocínio (Kosmyna *et al.*, 2025).

A longo prazo, essa falta de exercício mental pode enfraquecer certas habilidades cognitivas. A confiança cega nos sistemas de IA levanta a preocupação de que o julgamento humano, a experiência e a empatia sejam abandonadas em favor de soluções automatizadas, que podem ser falhas, tendenciosas ou até mesmo gerar informações falsas.

Apesar disso, ambos os estudos apontam que a solução não consiste em deixar de usar a IA, mas em adotar uma abordagem mais equilibrada. Em vez de permitir que a IA substitua o pensamento humano, a recomendação é usá-la como uma ferramenta para aprimorar e complementar nossas capacidades cognitivas.

Esse modelo híbrido de colaboração entre humanos e máquinas propõe que as decisões da IA devem ser sempre verificadas e avaliadas por um ser humano. O objetivo é garantir que a IA ajude a construir um futuro mais produtivo, sem nos tornar menos pensantes.

1.3 Espelho da Sociedade: a quem a IA representa?

A inteligência artificial pode ser compreendida como um espelho da sociedade que a cria. Longe de ser uma entidade neutra ou universal, ela carrega, em sua própria estrutura e funcionamento, as marcas culturais, linguísticas e cognitivas de seus criadores (Córdova, 2025). Cada modelo é, de certo modo, uma síntese do mundo que o alimenta — suas crenças, valores, modos de pensar e hierarquias de conhecimento.

Essa relação entre tecnologia e cultura é demonstrada de forma clara no artigo “Which Humans?” de Atari *et al.* (2023), que investiga quais traços humanos são, de fato, refletidos nos modelos de linguagem atuais. Os autores buscaram compreender de que forma a psicologia das GenAIs espelha a diversidade psicológica humana — ou, mais precisamente, quais aspectos dessa diversidade são reproduzidos e quais são deixados de fora.

Para isso, aplicaram um método comparativo, submetendo um modelo GPT, da OpenAI, a instrumentos clássicos de pesquisa das ciências sociais e cognitivas, e confrontando suas respostas com grandes bases de dados empíricos sobre valores e atitudes humanas. O principal referencial utilizado foi o *World Values Survey* (WVS), com mais de 94 mil participantes de 65 países, que oferece um retrato global das crenças e comportamentos culturais (Atari *et al.*, 2023). Os resultados desse estudo são discutidos a seguir.

1.3.1 O mito do “humano genérico”

Nas últimas décadas, a inteligência artificial — especialmente os Grandes Modelos de Linguagem (LLM²) — passou a ocupar um lugar central no debate sobre o que significa “pensar” e “compreender”. Esses sistemas, treinados com bilhões de palavras extraídas de fontes públicas, disponíveis principalmente na internet, geram textos,

² LLM: adotamos a sigla do nome em inglês *Large Language Models*, pelo qual são mais conhecidas.

respondem perguntas, redigem códigos e até formulam argumentos morais. Tanto que, frequentemente, estudos e relatórios técnicos descrevem essas tecnologias como possuidoras de “nível humano” de desempenho em diversas tarefas cognitivas (Bubeck et al., 2023).

Mas dizer que um modelo de IA pensa como um humano traz muitas questões que precisam ser debatidas. Dentre elas: o que significa pensar como um humano, e de quais humanos estamos falando? (Atari et al., 2023).

É precisamente sobre estas questões que o estudo de Atari et al. (2023) discute de maneira muito clara. O primeiro ponto do trabalho problematiza uma suposição implícita nas comparações entre IA e humanidade: a ideia de que existe uma psicologia universal, representada por um “humano médio”.

Entretanto, o estudo demonstra, com base empírica, que essa suposição é falsa — e que os grandes modelos de linguagem refletem, de modo marcante, apenas uma fração específica da diversidade humana: os chamados povos WEIRD (*Western, Educated, Industrialized, Rich, Democratic*).

A psicologia humana não é homogênea. Décadas de pesquisa em antropologia, psicologia cultural e ciências cognitivas têm mostrado que populações ao redor do mundo divergem amplamente em suas formas de pensar, raciocinar, cooperar, julgar moralmente e até perceber o mundo sensorial (Atari et al., 2023).

Sociedades ocidentais tendem a valorizar a individualidade, a autonomia e o raciocínio analítico. Outras culturas priorizam a

interdependência, as relações sociais e o contexto coletivo. Há variações marcantes na forma de lidar com autoridade, com normas morais, com o tempo e com a linguagem. Não há, portanto, um “modo humano” único de pensar, sentir e decidir — há uma multiplicidade de psicologias moldadas pela cultura (Atari *et al.*, 2023). E isso tudo precisa ser considerado no contexto educacional. O estudante é um sujeito com múltiplas dimensões: social, política, religiosa, afetiva etc. Não se pode desconsiderar isso.

Contudo, grande parte dos dados textuais disponíveis na internet — principal fonte de aprendizado dos LLMs — provém justamente de países ocidentais e anglófonos, onde o acesso à escrita, à tecnologia e à conectividade é mais amplo. Assim, as máquinas aprendem com um retrato parcial da humanidade.

1.3.2 Implicações da homogeneização do pensar: a bolha do conhecimento da IA

A consequência de treinar modelos de IA com dados majoritariamente produzidos por populações WEIRD é que esses modelos tendem a reproduzir o mesmo viés cultural, projetando sobre o mundo uma visão ocidentalizada de moralidade, racionalidade e identidade. Isso implica que, ao interagir com uma IA generativa, o usuário pode se deparar com uma psicologia artificialmente universalizada — uma voz que aparenta representar a humanidade, mas que, na prática, expressa apenas a perspectiva de uma fração privilegiada dela (Atari *et al.*, 2023).

Tal cenário tem implicações profundas em diferentes dimensões:

- **Dimensão científica:** os LLMs têm sido amplamente utilizados em pesquisas científicas e aplicações sociais, mas carregam um viés cultural significativo por serem treinados majoritariamente com dados de populações WEIRD. Esse viés compromete a representatividade dos modelos, pois suas respostas refletem valores, crenças e estilos cognitivos específicos dessas populações, ignorando a diversidade psicológica global.

Isso limita a validade de estudos que usam LLMs para inferir comportamentos humanos, já que os modelos não capturam adequadamente as variações culturais em julgamentos morais, preferências econômicas ou estilos de pensamento. Além disso, o crescimento do tamanho dos modelos não resolve esse problema, pois a diversidade dos dados continua sendo um fator crítico para a qualidade e equidade dos resultados (Atari et al., 2023).

- **Dimensão cultural e ética:** ao reproduzir normas WEIRD, os LLMs podem reforçar uma visão de mundo ocidental como padrão, marginalizando outras culturas e formas de raciocínio. Isso levanta preocupações sobre colonialismo digital, exclusão social e injustiça algorítmica, especialmente quando esses

modelos são aplicados em contextos sensíveis como educação, saúde ou políticas públicas (Atari *et al.*, 2023).

A falta de diversidade nos dados e nas equipes de desenvolvedores também compromete a capacidade dos modelos de compreender nuances locais, o que pode gerar respostas inadequadas ou ofensivas. Para mitigar esses riscos, é essencial que pesquisadores e desenvolvedores reconheçam esse fato e busquem ampliar a representatividade cultural nos dados de treinamento, respeitando as múltiplas definições de diversidade existentes ao redor do mundo (Atari *et al.*, 2023).

Trata-se, portanto, de um problema tanto tecnológico, quanto político. Não basta adicionar mais idiomas ou textos traduzidos — é preciso representar formas de pensar, valores e contextos diversos, reconhecendo que a pluralidade humana é parte essencial da inteligência. Isso implica promover acesso e inclusão. Desafios globais para alcançar o verdadeiro desenvolvimento.

1.4 Agentes de IA: uma nova fronteira

Além de tudo o que já foi discutido até aqui, podemos dizer que nos encontramos à beira de uma nova fronteira tecnológica: os agentes de IA. Agentes, nesse contexto, são sistemas de software capazes de executar tarefas de maneira relativamente autônoma e perseguir objetivos definidos pelo usuário sem depender de supervisão contínua. Esses sistemas interagem com o ambiente, analisam informações,

aprendem com suas próprias experiências e tomam decisões para realizar tarefas que podem ser complexas e multifacetadas.

O impacto dessa tecnologia no campo da educação reside na possibilidade de professores e estudantes criarem seus próprios agentes para realizar tarefas em seu nome. Atualmente, a maioria dos agentes que operam em um LLM, como o GPT, está restrita ao ambiente do navegador web. No entanto, o potencial desse recurso tecnológico sugere uma expansão para além desse limite, permitindo que o agente saia do navegador e obtenha acesso ao computador de cada usuário e, principalmente, aos aplicativos nele instalados.

Para ilustrar esse potencial, pode-se imaginar o seguinte cenário: a professora atribui um trabalho para ser realizado, em grupos, fora da sala de aula (dever de casa). O estudante **A** determina ao seu agente que vá resolver o trabalho de casa em colaboração com o agente do estudante **B**, que por sua vez também definiu a mesma tarefa para seu agente. O resultado é que a tarefa de casa acabará sendo executada pelos agentes dos estudantes **A** e **B**.

O potencial de delegação de tarefas de aprendizagem a agentes de IA, como ilustrado, configura um impacto disruptivo na dinâmica educacional. Essa capacidade de os agentes realizarem tarefas complexas e colaborativas levanta questões fundamentais sobre a natureza do aprendizado, a autenticidade da produção estudantil e a validação da avaliação.

A transição da execução humana para a automação por IA exige uma redefinição dos objetivos pedagógicos, focando nas habilidades

cognitivas de nível superior (como o *design* e a supervisão da tarefa do agente) em vez da mera execução, e impõe a necessidade urgente de desenvolver novos métodos de avaliação capazes de avaliar a compreensão e a capacidade de direcionamento do estudante, e não apenas o produto final gerado pela máquina.

Todas essas questões pedem a revisão de políticas no uso de tecnologias de IA. Por se tratar de problemas recentes e complexos, ainda não há consenso sobre a melhor solução para enfrentá-los. Entretanto, existem importantes discussões e orientações que podem ajudar nesse processo. Os próximos capítulos buscam construir uma base teórica sólida sobre como tratar dos problemas éticos envolvendo inteligência artificial.

2. ÉTICA E GOVERNANÇA: necessidade urgente

Definir o que é ética e até onde ela vai é um dos maiores desafios da humanidade. A ética trata das escolhas que fazemos sobre o que é certo ou errado, justo ou injusto — e essas escolhas mudam conforme o tempo, a cultura e as situações. Quando falamos de IA, tudo isso fica ainda mais desafiador. Estamos lidando com sistemas que aprendem, tomam decisões e, muitas vezes, influenciam diretamente a vida das pessoas.

Com o avanço da tecnologia, sistemas de IA — como *chatbots*, robôs, assistentes virtuais e sistemas de recomendação — deixaram de ser vistos apenas como ferramentas. Hoje, eles são como parceiros de trabalho, ajudantes ou até “colegas” em várias tarefas. Isso abriu um novo e importante campo de discussão: entender o impacto ético que essas tecnologias têm sobre a sociedade (Dignum, 2018).

Quando a IA começa a tomar decisões sozinha ou a influenciar o que os humanos decidem, surgem perguntas difíceis: quais são as consequências morais, legais e sociais dessas decisões? Um sistema pode ser responsabilizado pelos seus erros? Como garantir o controle sobre uma tecnologia que aprende e muda com o tempo? E até que ponto devemos permitir que esse tipo de inovação avance? (Córdova, 2024).

A forma como a sociedade vai lidar com questões como essas vai definir o quanto poderemos confiar e conviver de maneira segura com a IA. No fim das contas, agir de forma ética nesse campo significa tentar garantir que as decisões envolvendo inteligência artificial

respeitem a dignidade humana, promovam justiça e contribuam para o bem-estar de todos. Esses devem ser os pilares que sustentam o desenvolvimento e o uso responsável da inteligência artificial.

Nesse cenário, entra em cena a governança em IA, um elemento essencial para assegurar que essas tecnologias sejam usadas de forma responsável. A governança diz respeito a criar estruturas e práticas que orientem o desenvolvimento, o uso e a supervisão da IA com base em princípios éticos e valores humanos.

Ela envolve definir regras claras, responsabilidades e mecanismos de controle que assegurem transparência, justiça, privacidade e prestação de contas. Em outras palavras, é o que ajuda a manter a IA a serviço das pessoas — e não o contrário. Sem uma boa governança, corremos o risco de ver a tecnologia avançar mais rápido do que nossa capacidade de entender e lidar com seus impactos sociais, culturais e educacionais.

Para compreender melhor como lidar com esses desafios é importante estar ciente das dimensões impactadas por esses novos recursos tecnológicos e conhecer um pouco sobre os princípios que devem nortear o seu desenvolvimento. Para isso, o presente capítulo apresenta, a seguir, estas questões sob uma ótica mais generalista, que transcende o âmbito da Educação.

Isso porque o estudo da ética em IA é pensado para guiar o desenvolvimento e implantação de soluções que a implementam de forma ampla. Ainda assim, tudo o que for apresentado se aplica

também a domínios como ensino, aprendizagem, gestão escolar, e qualquer outro que possa se beneficiar dessa tecnologia.

2.1 Dimensões da Ética em IA

Falar sobre ética em IA é um desafio, porque nos leva a considerar tecnologias que aprendem, decidem e interagem de formas cada vez mais autônomas, e definir o que é “certo” ou “errado” nesse contexto não é algo simples. Para lidar com essa complexidade, pesquisadores e organismos internacionais têm proposto diferentes dimensões de abordagem, que ajudam a entender como princípios e valores podem ser incorporados ao ciclo de vida da IA.

Quanto às dimensões de abordagem, podemos considerar a ética por design, a ética em design e a ética para o design. Cada uma abrangendo um conjunto específico de questões que levam a procura de soluções para tornar a IA mais segura e confiável. Vejamos.

2.1.1 Ética em design (ethics in design)

A Ética em Design foca no processo de desenvolvimento e nas implicações sociais que a IA pode causar. Em vez de se concentrar nos bits e bytes do código, essa abordagem busca investigar questões do tipo: "Como esse sistema vai impactar a vida das pessoas? Que problemas sociais ele pode criar ou agravar?" Ela considera o contexto humano e social em que a tecnologia será inserida.

Com isso, a ética por design não se preocupa apenas se a tecnologia é justa, mas se ela é benéfica para a sociedade. Por exemplo, imagine que uma instituição de ensino decide usar sistemas tutores inteligentes para substituir professores em aulas de educação a distância (EaD).

A abordagem de Ética em Design não se perguntaria se a IA pode ensinar (isso seria uma questão técnica), mas sim se ela deve. A análise ética aqui consideraria as implicações sociais e pedagógicas:

- **Impacto social:** a substituição de professores desvalorizaria a profissão? O que acontece com a interação humana e o desenvolvimento de habilidades sociais que a presença de um professor proporciona?
- **Impacto na aprendizagem:** a IA pode adaptar o conteúdo, mas ela consegue identificar problemas emocionais ou dificuldades de aprendizado que não são óbvias? A IA pode oferecer o apoio e a empatia que um professor humano oferece?

A ética em design exigiria que a instituição considerasse os riscos e benefícios para estudantes, professores e para a qualidade da educação como um todo, antes de simplesmente adotar a tecnologia.

Outro exemplo clássico é o de uma IA usada para reconhecimento facial. Antes de implementar uma solução desse tipo, a ética em design perguntaria: Essa tecnologia vai ser usada para segurança ou para vigilância em massa? Ela poderia violar a

privacidade de cidadãos inocentes? E quanto às questões de viés racial? A discussão aqui acontece em torno do uso e do impacto social da tecnologia.

2.1.2 Ética para o design (ethics for design)

A Ética para o Design trata dos códigos de conduta e das responsabilidades dos profissionais que projetam, desenvolvem e usam sistemas de IA. É a ética do indivíduo, que incentiva a integridade, a transparência e o compromisso com o bem comum.

Para compreender melhor, imagine que uma empresa esteja criando um sistema de reconhecimento facial. Durante alguma etapa do processo, um desenvolvedor descobre que o sistema tem uma taxa de erro muito alta para pessoas de pele não branca.

Uma abordagem de Ética para o Design exigiria que o desenvolvedor tenha a responsabilidade de comunicar esse problema à sua equipe e, se necessário, recusar o lançamento do produto até que o viés seja corrigido, mesmo sob pressão.

Essa ética também se aplica à transparência. Se um engenheiro de IA usou dados de uma forma que poderia ser mal interpretada ou prejudicial, a ética profissional o obriga a divulgar essa informação de maneira clara e honesta em relatórios ou publicações.

2.1.3 Ética por design (*ethics by design*)

A Ética por Design é a abordagem mais profunda, pois ela busca incorporar a ética diretamente na arquitetura técnica dos sistemas de IA. A ideia principal é que a ética não seja apenas uma camada superficial ou uma reflexão tardia, mas sim um pilar fundamental da tecnologia, assim como a segurança ou a usabilidade. Trata-se do desafio de codificar valores humanos em algoritmos que possibilitem às máquinas, tomar decisões que possam ser consideradas éticas.

Tomemos novamente o exemplo de um STI que personaliza o aprendizado para cada estudante. Uma abordagem de ética por design, nesse caso, não focaria apenas em otimizar o desempenho, mas também em proteger a autonomia do estudante.

O algoritmo seria projetado com uma regra interna para evitar a intervenção excessiva. Em vez de corrigir o estudante a cada erro, o sistema esperaria que ele tentasse resolver o problema por conta própria em um número razoável de vezes. Se o estudante persistir no erro, a intervenção do sistema seria gradual e suave, oferecendo uma dica em vez da resposta completa.

Em carros autônomos, a ética por design se torna extremamente relevante, pois esses sistemas precisam tomar decisões críticas em situações de risco. Imagine, por exemplo, um cenário em que o veículo detecta um pedestre atravessando a rua de forma inesperada. O carro precisaria decidir entre desviar do caminho, evitando o atropelamento do pedestre, mas colocando o próprio

motorista em risco, ou manter a rota para proteger o motorista, com o risco de ferir o pedestre.

Nesses casos, os algoritmos não estão apenas executando comandos mecânicos; eles estão diante de escolhas que envolvem valores éticos complexos. A programação de ética por design busca antecipar essas situações, definindo critérios claros que equilibrem segurança, justiça e responsabilidade, para que as decisões tomadas pelo veículo reflitam princípios éticos previamente estabelecidos.

2.2 Princípios para Ética em IA

Nos debates globais sobre o uso ético da IA, diversas organizações têm assumido um papel central na definição de diretrizes capazes de orientar seu desenvolvimento responsável. Entre elas, destacam-se a UNESCO, agência das Nações Unidas dedicada à educação, ciência e cultura, o *Instituto de Engenheiros Eletricistas e Eletrônicos* (IEEE), maior organização profissional do mundo nas áreas de tecnologia e engenharia, a Organização para a Cooperação e Desenvolvimento Econômico (OCDE), um fórum internacional onde países-membros discutem e promovem políticas públicas e padrões internacionais para o bem-estar socioeconômico global, entre outras. Todas elas propõem princípios que buscam assegurar que a IA seja projetada e utilizada em conformidade com valores humanos, sociais e éticos.

A seguir, são apresentados alguns desses princípios fundamentais.

- **Justiça e Equidade:** este princípio procura garantir que a IA atue de forma justa, evitando reproduzir ou intensificar desigualdades já existentes. Isso envolve o cuidado em identificar e corrigir vieses nos dados e algoritmos, que podem levar a discriminações de gênero, raça, idade, origem ou condição socioeconômica. Ao incorporar esses valores, busca-se que a IA contribua ativamente — ou, no mínimo, não dificulte — a redução das desigualdades e a ampliação do acesso a oportunidades. Dessa forma, a tecnologia se mantém a serviço das pessoas, promovendo inclusão, justiça social e respeito à diversidade humana.

Além disso, a equidade também diz respeito às condições de acesso às diferentes tecnologias disponíveis. Isso inclui questões como conectividade, recursos de hardware, software e mecanismos de inclusão.

- **Transparência e Explicabilidade:** no campo da ética em Inteligência Artificial, a transparência e a explicabilidade são fundamentais para construir confiança e responsabilidade no uso dessas tecnologias.

Transparência, nesse contexto, refere-se à clareza sobre como um sistema funciona e como suas decisões são produzidas. Isso envolve prover clareza sobre os dados utilizados, os critérios adotados e os processos internos que levam um

algoritmo a um resultado, permitindo que usuários, desenvolvedores e reguladores compreendam a lógica subjacente. A transparência busca criar confiança e possibilitar auditoria e supervisão, permitindo que a IA opere de maneira responsável e alinhada a princípios éticos.

Explicabilidade, por sua vez, diz respeito à capacidade de um sistema de IA justificar suas decisões de forma compreensível para os humanos. Enquanto a transparência foca em tornar o funcionamento do sistema acessível, a explicabilidade traduz essa complexidade em explicações claras sobre porque uma determinada recomendação ou decisão foi tomada.

Em contextos educacionais, a transparência e a explicabilidade permitem que professores e estudantes compreendam e questionem os resultados produzidos pela IA, fortalecendo o pensamento crítico e possibilitando decisões mais conscientes. Isso se torna ainda mais relevante quando tais sistemas são utilizados para recomendar conteúdos, avaliar desempenho ou personalizar o ensino, situações em que suas decisões podem influenciar diretamente o aprendizado.

Quando a lógica de funcionamento desses sistemas é opaca ou difícil de entender, há o risco de que decisões automatizadas sejam aceitas sem questionamento, o que pode comprometer a autonomia do educador e a equidade no processo de aprendizagem. Por isso, promover transparência não é apenas uma exigência técnica: é um compromisso ético com a clareza,

a responsabilidade e o direito de entender como a tecnologia impacta as pessoas.

- **Privacidade e Proteção de Dados:** esse princípio abrange o conjunto de normas, práticas e tecnologias destinadas a resguardar as informações de indivíduos, garantindo o direito fundamental de cada pessoa ao controle sobre seus próprios dados pessoais, conforme previsto na Lei Geral de Proteção de Dados (Brasil, 2018, Art. 1º e 2º).

De modo geral, o foco está em estabelecer que a coleta, o armazenamento, o processamento e o compartilhamento de qualquer dado — como nome, endereço, e-mail ou informações sensíveis — devem obedecer aos princípios de finalidade, adequação, necessidade, transparência e segurança (Brasil, 2018, Art. 6º).

No contexto educacional, essa responsabilidade adquire uma camada especial de cuidado, pois muitas vezes envolve dados de crianças e adolescentes, considerados sujeitos em condição peculiar de desenvolvimento e titulares de direitos fundamentais, conforme estabelece o Estatuto da Criança e do Adolescente (Brasil, 1990, art. 2º e 3º). Com isso, proteger as informações dos estudantes — como notas, frequência, históricos comportamentais, dados de saúde e registros familiares — é fundamental.

As instituições de ensino devem ser as guardiãs desses ativos, garantindo que o uso dos dados seja seguro, consentido

(especialmente dos pais ou responsáveis para menores de idade) e com finalidades pedagógicas legítimas.

- **Supervisão Humana:** um dos pilares mais importantes para uma implementação ética e eficaz da inteligência artificial no ambiente de trabalho ou educacional reside na garantia da supervisão humana. Este princípio estabelece que professores, estudantes, gestores e demais profissionais devem ser mantidos firmemente no centro do processo decisório.

A IA deve ser encarada estritamente como tecnologia de apoio, aceleração e suporte à análise, e não como um substituto autônomo para o julgamento, a crítica ou a decisão final. O objetivo é que a tecnologia aumente a capacidade humana, automatizando tarefas e fornecendo *insights* baseados em dados, mas que a responsabilidade e a autoridade permaneçam inegavelmente com o ser humano. Dessa forma, é mais fácil assegurar a ética, a empatia e a contextualização necessárias que os sistemas algorítmicos não podem oferecer.

- **Prestação de Contas (Accountability):** quem deve responder se uma decisão automatizada tomada por IA causar danos a pessoas, ao meio ambiente ou ecossistemas? E no âmbito escolar, a quem cabe essa responsabilidade? É buscando garantir uma resposta para questões como esta que foi proposto o princípio da prestação de contas.

Este princípio se refere à necessidade de assegurar que todas as partes envolvidas no desenvolvimento, implantação e uso de

sistemas de IA — como desenvolvedores, gestores e instituições — tenham responsabilidades claramente definidas e possam responder por decisões e ações tomadas por recursos de inteligência artificial. Isso implica criar processos de registro e acompanhamento das etapas do desenvolvimento, bem como mecanismos de auditoria que permitam identificar a origem de erros, vieses ou impactos negativos causados pela tecnologia.

Além disso, envolve a implementação de procedimentos de correção e de comunicação transparente com os usuários e com a sociedade, assegurando que eventuais falhas sejam reconhecidas, explicadas e tratadas de maneira ética e responsável.

- **Letramento em IA (AI Literacy):** consiste em desenvolver a capacidade de compreender de forma crítica o que é a Inteligência Artificial, como ela funciona e quais são seus limites e potenciais impactos éticos, sociais e culturais. Essa competência vai além do simples uso de ferramentas automatizadas: envolve entender os princípios que orientam o funcionamento dos algoritmos, como eles processam dados e tomam decisões, e de que maneira podem reproduzir vieses ou gerar consequências imprevistas.

Ao promover essa compreensão, educadores e estudantes tornam-se aptos a identificar riscos, questionar resultados e utilizar a tecnologia de modo mais informado e responsável. O

letramento em IA, portanto, é fundamental para que a sociedade não apenas consuma tecnologias inteligentes, mas também participe ativamente da discussão sobre sua aplicação, regulação e desenvolvimento.

Por fim, discutir ética e governança em IA é, acima de tudo, refletir sobre como equilibrar inovação e responsabilidade — garantindo que a tecnologia sirva ao desenvolvimento humano, à equidade e à formação de cidadãos críticos, e não apenas à eficiência, ao lucro ou a interesses políticos e corporativistas.

3. USO ÉTICO DA IA NA PRODUÇÃO DE CONTEÚDO

A integração da GenAI na produção de conteúdo representa uma das transformações mais significativas na sociedade e no ambiente acadêmico e corporativo moderno. A capacidade de gerar rapidamente textos, vídeos, apresentações, imagens, áudio etc., se bem usados, pode impulsionar a eficiência e a criatividade e democratizar o acesso a recursos que antes demandavam conhecimento especializado (Córdova, 2025). No entanto, essa velocidade e facilidade impõem um conjunto complexo de desafios éticos, regulatórios e autorais.

Nesse cenário, é preciso, primeiramente, destacar que a solução não deve incluir a proibição do uso, mas sim, o estabelecimento de protocolos rigorosos para garantir que a aplicação desses recursos tecnológicos mantenha a credibilidade, a responsabilidade e, acima de tudo, a autoria intelectual humana, que são os pilares inegociáveis do trabalho científico e profissional. O foco precisa ser deslocado da capacidade técnica da IA para a responsabilidade fundamentalmente do usuário humano.

Assim, os próximos capítulos oferecem reflexões e boas práticas para o uso e aplicação de tecnologias de inteligência artificial generativa na educação e na produção científica. O objetivo não é servir como um guia definitivo, mas fomentar uma discussão sobre esse tema tão presente na vida de professores e pesquisadores.

3.1. A Erosão da Autoria: entre a produtividade e a responsabilidade intelectual

O uso ético da IA no âmbito da pesquisa e da produção de conteúdo, atualmente, é fortemente determinado pelo princípio de maximizar os benefícios de produtividade — como a sumarização eficiente de grandes volumes de dados e a melhoria da clareza linguística — enquanto se busca mitigar os riscos inerentes à tecnologia ou ao seu mau uso. Nesse contexto, conforme já demonstrado até aqui, entre os principais riscos estão a produção de imprecisões factuais (alucinações), a superficialidade e falta de contexto, o reforço de preconceitos discriminatórios e a erosão da autoria humana.

Em meio a esse dilema, um estudo publicado em 2024 mostra um aumento significativo no uso de inteligência artificial para a escrita de artigos científicos (Haider, 2024). Isso levanta preocupações e dúvidas sobre a integridade da ciência e deixa clara a necessidade de ações para mitigar esse fenômeno.

Contudo, é fundamental reconhecer que parte da solução consiste em compreender a GenAI como uma tecnologia de apoio — e não como uma entidade verdadeiramente inteligente, capaz de resolver todos os problemas de produtividade humana. Ferramentas impulsionadas por IA podem auxiliar eficientemente em tarefas como pesquisa preliminar, identificação de literatura relevante, geração de citações e refinamento da legibilidade do manuscrito.

O problema surge quando esses recursos são utilizados para produzir conteúdo de pesquisa original, posteriormente atribuído ao pesquisador, ou quando se delega a eles a elaboração de textos sem uma revisão cuidadosa ou sem critérios humanos claramente estabelecidos. Esse tipo de uso é antiético e acarreta sérios riscos, pois a IA carece do pensamento crítico, do raciocínio analítico e da expertise de área, necessários para o trabalho intelectual legítimo.

A distinção entre assistência e autoria é reforçada por um consenso rigoroso entre editoras científicas de alto impacto como *Elsevier*, *Taylor & Francis*, *The Wiley* e *Springer Nature*, universidades e órgãos de propriedade intelectual, como o *World Intellectual Property Organization* (WIPO). Estes órgãos proíbem consistentemente que a IA seja listada como autora ou coautora de um trabalho.

A autoria implica concepção intelectual, e a capacidade de assumir responsabilidade pelo conteúdo publicado configura-se como um atributo exclusivamente humano. No entanto, é preciso reconhecer que o conceito de originalidade estrita é complexo e passível de questionamento, dada a inevitável influência de trabalhos prévios e do conhecimento acumulado no processo de geração de conteúdo.

Apesar disso, o protagonismo e a intervenção humana são indispensáveis para garantir a precisão, a síntese e a adequação contextual do material. Consequentemente, cabe ao pesquisador manter o controle autoral e a responsabilidade total e inalienável sobre o manuscrito final.

Em última instância, o desafio ético não está apenas em usar a IA corretamente, mas em não permitir que ela nos substitua naquilo que é humano e que define o próprio ato de pesquisar: o pensar. A tentação de terceirizar o raciocínio à máquina é sutil, mas perigosa — especialmente quando os resultados parecem tão convincentes. O risco é que a ciência, em busca de eficiência, acabe se tornando reprodutiva em vez de criativa, repetindo padrões gerados por algoritmos em vez de expandir fronteiras do conhecimento.

A IA, portanto, não ameaça apenas o autor; ameaça a própria noção de autoria como gesto de invenção e de responsabilidade. Nesse novo ecossistema intelectual, é preciso reforçar que a ética não deve ser vista como um freio, mas como um ato de resistência cognitiva.

Usar a IA com consciência e propósito claro significa reivindicar o papel do humano como intérprete, crítico e curador do conhecimento. Não se trata de rejeitar a tecnologia, mas de reafirmar a centralidade da mente humana diante de uma era que parece preferir a velocidade à reflexão.

3.1.1 Riscos éticos centrais da IA generativa

Uma ironia curiosa sobre o uso irresponsável da GenAI é que este muitas vezes decorre justamente das limitações presentes nos próprios modelos que viabilizam esta tecnologia — os LLMs e os modelos multimodais. E ainda que tais limitações possam ser oriundas do processo de treinamento — como dados enviesados, escolhas

algorítmicas ou decisões de projeto —, fazendo parecer que o problema é puramente técnico, a raiz da questão está no desconhecimento generalizado sobre como essas ferramentas funcionam e do que são realmente capazes.

Parte dessa incompreensão vem das terminologias que usamos para nos referirmos às tecnologias de IA. Palavras como “inteligência”, “aprendizado”, “raciocínio” e “criatividade”, fazem parecer que as máquinas estão realmente desenvolvendo capacidades humanas. Isso gera um ruído conceitual que confunde e contribui para a mitificação das técnicas de computação e da própria IA.

Chama-se de abordagem antropomórfica o ato de atribuir características humanas a animais ou coisas. Mas ao fazer isso com a IA, apesar de facilitar a comunicação e a compreensão de alguns fenômenos, corre-se o risco de distorcer a compreensão pública sobre o que essas tecnologias realmente são (Watson, 2019; Salles *et al.*, 2020).

No meio acadêmico, essa antropomorfização é aceita com a devida cautela: entende-se que são metáforas úteis para descrever funções técnicas, não capacidades humanas no sentido estrito. Assim, combater os efeitos da antropomorfização mal compreendida é parte essencial do desafio. Daí o papel do letramento em IA.

Para isso, o primeiro passo é compreender que ferramentas que usam recursos de IA não têm, de fato, a noção da realidade subjacente, nem raciocínio causal e menos ainda consciência daquilo que a linguagem descreve, focando apenas em satisfazer a consistência

estatística com o prompt do usuário (Saba, 2024; Córdova, 2025). Isso explica muito sobre o porquê de ser tão crítico quando uma GenAI tenta gerar informações factuais, chegando a fazer citações que não existem.

Outro ponto importante a considerar é que, ao contrário da percepção comum, tecnologias de GenAI também não são genuinamente criativas. Os resultados que elas produzem são estritamente baseados em um vasto conjunto de informações conhecidas. Nesse sentido, a IA generativa, assim como as formas de IA previamente desenvolvidas, é fundamentalmente preditiva: sua capacidade se restringe a gerar resultados através da recombinação e extrapolação de padrões existentes nos dados históricos, não sendo capaz de conceber cenários ou conceitos verdadeiramente imaginativos e inéditos.

Mas além das limitações cognitivas das GenAIs, há outro aspecto igualmente preocupante: o modo como elas refletem — e frequentemente amplificam — os padrões existentes em seus dados de treinamento. Se, por um lado, a ausência de compreensão real do mundo leva à geração de informações imprecisas, por outro, a dependência de grandes volumes de dados coletados de contextos históricos e culturais diversos expõe essas ferramentas aos vieses neles contidos (Córdova, 2025).

Essa combinação entre falta de discernimento e aprendizado estatístico cego cria terreno fértil para a reprodução de distorções e injustiças sociais. Os dados de treinamento em massa frequentemente contêm preconceitos sociais, raciais ou de gênero. Quando não se

toma os devidos cuidados ao fazer uso de modelos treinados com tais dados, corre-se o risco de estar reforçando e perpetuando os mesmos preconceitos.

Na educação e na ciência isso se torna particularmente preocupante, pois ao institucionalizar o uso de ferramentas com essas características, pode-se também estar institucionalizando comportamentos discriminatórios. Um exemplo marcante desse tipo de falha foi a controvérsia envolvendo o Gemini, quando a IA, ao tentar corrigir vieses históricos, acabou produzindo imagens distorcidas e historicamente imprecisas (Ananya, 2024). Este caso evidenciou o quão desafiador é auditar e controlar a forma como a representatividade é tratada nos conteúdos gerados por inteligência artificial.

Por fim, a classificação de alto risco atribuída à aplicação de sistemas de IA no ambiente educacional, conforme prevista no Projeto de Lei nº 2.338/2023, que regula a inteligência artificial no Brasil, evidencia a necessidade de uma atenção especial a essas tecnologias. O PL reconhece que o uso de IA em contextos educativos envolve potenciais impactos significativos, exigindo que educadores e pesquisadores adotem práticas responsáveis e críticas na sua utilização.

Além disso, em alinhamento com os princípios discutidos no capítulo “2.2 Princípios para Ética em IA”, o projeto reforça que sistemas de IA aplicados à educação devem contar com supervisão humana constante, transparência nos processos e mecanismos claros de

responsabilização por possíveis danos. Como as máquinas carecem de raciocínio analítico e de agência moral, a ética no uso da IA depende diretamente do nível de intervenção, monitoramento e controle exercido pelo usuário humano.

3.2 Ferramentas Generativas como Apoio à Produção: limites e possibilidades

O uso responsável da GenAI exige a definição de critérios claros, que dependem do tipo de conteúdo a ser produzido. Os LLMs são considerados modelos de propósito geral, pois podem auxiliar de diferentes formas e em diferentes níveis no processo de produção de conteúdo.

Para a produção de textos e documentos acadêmicos, a GenAI oferece possibilidades significativas de otimização e produtividade. A IA é melhor que os humanos em tarefas de grande volume e repetição (Maurya; Maurya, 2024), incluindo a revisão e organização de referências bibliográficas e a análise e resumo de grandes volumes de literatura científica. Além disso, o uso da IA para fazer melhorias na gramática e tornar o texto mais claro e fluído é legítimo, desde que o controle autoral seja mantido pelo pesquisador.

Para demonstrar como a GenAI pode realmente ajudar, um estudo de 2024 analisou 100 artigos acadêmicos, sendo 50 produzidos por humanos e 50 gerados por modelos GPT, com o objetivo de comparar a confiabilidade, a precisão factual e a coerência textual

entre ambos. A pesquisa utilizou uma metodologia mista, combinando análise quantitativa — para medir erros de citação, veracidade das informações e consistência textual — com uma avaliação qualitativa conduzida por especialistas de diferentes áreas, que examinaram profundidade, clareza e valor acadêmico do conteúdo.

Os resultados mostraram que a inteligência artificial se destacou em fluência linguística e estrutura textual, produzindo textos coerentes e bem organizados, com desempenho ligeiramente superior ao humano nesses aspectos. Contudo, apresentou graves limitações em precisão factual, contextualização e qualidade das citações, alcançando um índice de 55% de erros em referências e 40% de informações incorretas (Maurya; Maurya, 2024).

Por outro lado, os autores humanos superaram a IA em pensamento crítico, profundidade conceitual e rigor metodológico, preservando a originalidade e a confiabilidade científica. O estudo concluiu, conforme discutido até agora, que o uso da IA pode ser vantajoso como ferramenta de apoio em tarefas repetitivas e de pré-redação, mas a autoria e a validação final devem permanecer sob responsabilidade humana, garantindo a integridade ética e intelectual da produção acadêmica.

Mas as possibilidades de uso da GenAI vão ainda além. Em 2024 a Unesco publicou um guia com “Recomendações sobre a Inteligência Artificial Generativa na Educação e na Pesquisa” (Unesco, 2024). Esse guia apresenta um conjunto de sugestões e cuidados no uso da IA para

estes fins. A seguir é mostrado um resumo dessas recomendações e os riscos associados:

- **Geração de conteúdo multimodal:** a GenAI pode criar textos, imagens, vídeos, músicas e códigos, automatizando a produção de materiais educacionais, didáticos e científicos.

Risco: geração de informações falsas ou imprecisas (“alucinações”), além da possibilidade de violar direitos autorais ao reproduzir conteúdo sem citação adequada.

- **Apoio à pesquisa científica:** a GenAI auxilia na formulação de hipóteses, revisão de literatura e estruturação de projetos de pesquisa.

Risco: falta de verificação factual e uso de referências inexistentes, comprometendo a confiabilidade e a integridade do conhecimento produzido.

- **Suporte ao ensino e planejamento pedagógico:** professores podem utilizá-la para criar planos ensino, planos de aula, atividades, materiais didáticos, avaliações e materiais de apoio como: mapas mentais, áudios, vídeos, apresentações entre outros.

Risco: homogeneização de práticas pedagógicas e perda da autoria e originalidade docente, reprodução de vieses que podem levar ao reforço de desigualdades em materiais e

conteúdo didático, redução da diversidade de opiniões e ideias, além de potencial dependência excessiva das sugestões da IA.

- **Tutoria personalizada e aprendizagem adaptativa:** a GenAI pode atuar como tutora virtual, adaptando o conteúdo ao ritmo e às necessidades do estudante.
- **Risco:** viés algorítmico que pode reforçar desigualdades, ameaça a privacidade, além do risco de reduzir a interação humana, essencial para o desenvolvimento crítico e emocional.
- **Apoio à inclusão educacional:** a GenAI pode ajudar na criação de planejamentos e materiais acessíveis para estudantes com deficiências visuais, auditivas ou cognitivas.
Risco: ausência de contextualização cultural e linguística, o que pode gerar representações inadequadas ou discriminatórias.
- **Facilitação da aprendizagem baseada em projetos e pesquisas:** as tecnologias de GenAI podem auxiliar estudantes no planejamento, organização e revisão de trabalhos investigativos.
- **Risco:** uso superficial da IA, levando à reprodução mecânica de ideias e enfraquecendo a capacidade de análise crítica e autoria dos estudantes.

- **Inovação e experimentação pedagógica:** a GenAI pode apoiar práticas educacionais criativas e colaborativas, explorando novas metodologias de ensino como: sala de aula invertida, aprendizagem baseada em problemas, gameificação, entre outras.

Risco: experimentação sem controle ético ou institucional, podendo gerar impactos negativos na aprendizagem e na confiabilidade dos resultados educacionais.

Em síntese, a UNESCO enfatiza, uma vez mais, que o potencial da GenAI só se concretiza quando há supervisão humana contínua, transparência nos processos e responsabilidade ética. A ausência desses fatores transforma a inovação em risco, comprometendo a confiabilidade, a equidade e o papel formativo da educação.

3.2.1 Por onde começar?

Com a rápida expansão e diversificação de recursos de GenAI que podem ser usados por professores, pesquisadores e estudantes, surgem questões importantes sobre como integrá-los de forma efetiva no fluxo de trabalho. Quais são as principais ferramentas disponíveis para uso? Qual é a finalidade e o valor científico de cada uma? Como obter acesso e começar a utilizá-las? Estas são algumas das dúvidas mais comuns.

Pensando nisso, este capítulo tem como objetivo apresentar um panorama dos principais recursos de IA para pesquisadores e professores disponíveis atualmente. Não se trata de um guia definitivo, pois tecnologias de IA são melhoradas e ganham novas funcionalidades muito rapidamente. Além disso, novas plataformas surgem a cada ano. Contudo, a lista a seguir serve como uma orientação inicial para aqueles que desejam conhecer melhor este universo.

Assim, a seguir são apresentadas as principais plataformas de GenAI organizadas por sua finalidade, abrangendo funções essenciais como a busca estratégica por artigos, a avaliação da influência e credibilidade de publicações, a simplificação da compreensão de textos densos e o suporte na elaboração textual.

a. Busca e Descoberta:

Nesta seção, são apresentadas ferramentas que podem ser utilizadas para a busca e descoberta estratégica de artigos científicos. Os recursos destacados exploram bases de dados especializadas e reconhecidas pela comunidade acadêmica, o que lhes confere maior credibilidade e rigor metodológico. Por essa razão, são considerados mais confiáveis do que suas congêneres que realizam buscas na internet sem critérios acadêmicos definidos, muitas vezes retornando resultados superficiais ou de baixa qualidade. Além disso, tais ferramentas oferecem funcionalidades adicionais, como filtros avançados, indicadores de impacto e opções de exportação de

referências, que contribuem para uma pesquisa mais organizada, precisa e alinhada às exigências de trabalhos científicos e acadêmicos.

Tabela 2. Ferramentas para Busca e Descoberta.

Ferramenta	Link de Acesso	Finalidade Científica Detalhada
Consensus	https://consensus.us.app/search/	Possibilita a realização de buscas diretamente em bases científicas. Sua principal função é encontrar respostas baseadas em evidências e sintetizar o consenso científico sobre uma pergunta específica.
Elicit	https://elicit.com/	Auxilia na identificação e revisão de literatura científica. Permite fazer perguntas de pesquisa e extrai informações-chave (como métodos, amostras e resultados) dos artigos, organizando-as de forma estruturada para revisões sistemáticas ou rápidas.
Scispace	https://scispace.com/	Pode auxiliar em várias etapas de uma pesquisa acadêmica: realizar buscas abrangentes de literatura científica, analisar e comparar documentos PDF, extrair informações de artigos, ajudar em revisões sistemáticas, criar apresentações e pôsteres, gerar visualizações e análises de dados, buscar <i>journals</i> e conferências relevantes para publicação e encontrar editais de financiamento para pesquisa.
ResearchRabbit	https://researchrabbitapp.com/	Facilita a descoberta de literatura relevante ao mapear visualmente conexões entre artigos, autores e tópicos. É excelente para expandir referências, encontrar trabalhos similares e acompanhar novos desenvolvimentos em uma área.

Fonte: Elaborado pelos autores.

b. Análise de Artigos e Credibilidade:

A seguir, são apresentadas ferramentas de IAGen desenvolvidas para apoiar pesquisadores na análise, avaliação da influência e credibilidade de publicações científicas. Esses recursos permitem não apenas avaliar a influência de um artigo por meio de métricas como citações, fator de impacto e relevância em determinada área, mas também verificar a credibilidade das fontes utilizadas e a consistência metodológica dos estudos.

Tabela 3. Ferramentas para Análise de Credibilidade de Artigos.

Ferramenta	Link de Acesso	Finalidade Científica Detalhada
Scite	https://scite.ai/	Ferramenta voltada à análise de citações, que verifica a credibilidade de um artigo a partir da forma como ele é citado em outros trabalhos — se é apoiado, contestado ou apenas mencionado. Essa avaliação auxilia o pesquisador a construir argumentos mais consistentes, considerando como o texto foi recebido pela comunidade científica.
Evidence Hunt	https://evidenc ehunt.com/	Ferramenta projetada para identificar rapidamente evidências científicas que respondam a perguntas específicas. É muito útil para revisões rápidas e para apoiar a tomada de decisões na área da saúde baseada em dados concretos.

Fonte: Elaborado pelos autores.

c. Leitura e Compreensão de Artigos:

A seguir, são apresentadas ferramentas de IAGen capazes de auxiliar na leitura e compreensão de textos, oferecendo explicações simplificadas, resumos instantâneos e interpretações acessíveis de conceitos técnicos, equações matemáticas, tabelas e gráficos. Além disso, possibilitam a realização de perguntas sobre o conteúdo, promovendo uma leitura mais dinâmica e direcionada às necessidades específicas do estudo.

Tabela 4. Ferramentas para Estudo e Compreensão.

Ferramenta	Link de Acesso	Finalidade Científica Detalhada
ChatPDF	https://www.chatpdf.com/pt	Permite ao usuário fazer perguntas diretamente sobre o conteúdo de um PDF (artigo, relatório, livro). A IA lê e “compreende” o texto, quadros, tabelas e, em alguns casos, imagens, fornecendo resumos e respostas específicas para otimizar a leitura e a extração de dados.
ExplainPaper	https://www.explainpaper.com	Focado em simplificar artigos científicos complexos. O usuário pode selecionar qualquer trecho do texto para receber uma explicação imediata. Possibilita gerar explicações em diferentes níveis de complexidade.

Fonte: Elaborado pelos autores.

d. Geração, Estrutura e Revisão de Conteúdo:

A seguir são apresentadas ferramentas de IAGen para análise de texto e geração de conteúdo multimodal, ou seja, em diferentes formatos como texto, áudio, vídeo, imagem etc.

Tabela 5. Ferramentas para Geração, Estrutura e Revisão de Texto.

Ferramenta	Finalidade Científica Detalhada
ChatGPT, Copilot, Perplexity, Deepseek, Gemini, Notebooklm, Qwen, Claude, Gamma.	Ferramentas multimodais que permitem ao usuário fazer perguntas diretamente sobre o conteúdo de um PDF (artigo, relatório, livro). A IA lê e “compreende” o texto, quadros, tabelas e, em alguns casos, imagens, fornecendo resumos e respostas específicas para otimizar a leitura e a extração de dados. Algumas delas, como o Notebooklm podem gerar áudios, vídeos, mapas mentais, apresentações e planos de estudo.

Fonte: Elaborado pelos autores.

e. Visualização de Dados e Conceitos:

A seguir são apresentadas ferramentas de IAGen para geração de conteúdo conceitual.

Tabela 6. Ferramentas para Análise de Credibilidade de Artigos.

Ferramenta	Link de Acesso	Finalidade Científica Detalhada
Napkin	https://www.napkin.ai/	Tem como função principal transformar textos selecionados em recursos visuais (como imagens, mapas conceituais ou infográficos). É útil para criar anotações visuais e conectar conceitos de maneira mais intuitiva.

Fonte: Elaborado pelos autores.

3.3 Conteúdos Audiovisuais: considerações específicas

No campo da produção de conteúdo visual, a IAGen democratiza o acesso à expressão artística, permitindo o apoio à criação rápida de designs, ilustrações e visualizações conceituais para apresentações e outros tipos de materiais didáticos, sem a necessidade de treinamento técnico aprofundado.

Os riscos, nesse contexto, se concentram na autenticidade e na manipulação. Editoras científicas como Elsevier, Taylor & Francis e Springer Nature já possuem políticas estritas que proíbem o uso de IA para criar ou manipular dados e imagens que serão apresentados como resultados de pesquisa. A exceção é apenas para pequenas alterações como brilho, contraste e equilíbrio de cores ou criação de imagens baseadas em dados científicos que possam ser devidamente verificados e revisados (Springer Nature Limited, 2025; Elsevier, 2025; Taylor & Francis, 2025).

Além disso, a capacidade da IA de gerar conteúdo visual e de áudio convincente (*deepfakes*) representa um risco ético e social severo, especialmente no contexto eleitoral, legal e acadêmico, ameaçando a defesa da democracia e da verdade factual. Soma-se a isso os problemas de viés discriminatório presente em alguns modelos, que se não forem mitigados, podem promover a projeção de estereótipos igualmente discriminatórios.

Recentemente foram noticiados casos de artigos com erros graves em imagens geradas por IA. Em um dos casos constava uma

radiografia de perna humana com mais ossos do que o normal. No outro, uma imagem de rato com genitais desproporcionalmente grandes. Ambos os artigos precisaram ser retratados.

Atualmente, conforme mostrado no capítulo “3.2.1 Por onde começar?”, é bastante simples encontrar ferramentas que geram podcasts, vídeos e músicas de maneira muito facilitada para o usuário. Essas ferramentas podem impulsionar a criação de materiais didáticos de qualidade, mas precisam ser usadas de forma ética: com criação e supervisão humana, transparência e responsabilidade.

A transparência, nesse contexto, consiste em deixar explícito o uso de IA para apoio na geração do conteúdo presente no resultado final de um trabalho acadêmico (artigo, relatório, apostila, apresentação etc). Quem os consome precisa saber se são gerados por humano ou por inteligência artificial. Isso faz parte do compromisso ético que professores, pesquisadores e estudantes precisam assumir. O próximo capítulo abordará melhor como e quando fazer essa declaração de uso de IA.

4 TRANSPARÊNCIA E AUTORIA: um caminho a construir

Um dos pilares da transparência no uso da inteligência artificial consiste em deixar claro quando e como ela foi usada em um trabalho, algo que se faz por meio de uma *declaração de uso* (também conhecida como *disclosure*). Essa prática é essencial para promover a honestidade intelectual, pois permite que quem lê ou avalia o trabalho saiba até que ponto o conteúdo foi produzido por uma pessoa e quanto contou com a ajuda da IA. Assim, garante-se que o autor humano continue sendo responsável pelo que é apresentado e que o uso da tecnologia não esconda nem substitua a contribuição real do pesquisador.

4.1 Mandato de Transparência

A legislação de direitos autorais sempre se baseou na ideia de criação humana. No Brasil, a Lei de Patentes (Lei nº 9.279/1996) determina que apenas pessoas podem ser reconhecidas como autoras, mesmo quando a invenção ou o conteúdo tenha sido desenvolvido com auxílio parcial ou total de sistemas de inteligência artificial. Isso significa que produções geradas inteiramente por máquinas, sem participação criativa humana, não podem receber proteção autoral, reforçando o princípio ético de que o pesquisador ou professor deve ser o verdadeiro autor e responsável intelectual pelo trabalho.

Nesse sentido, a transparência sobre o uso dessas tecnologias ajuda a equilibrar inovação assistida por IA com responsabilidade intelectual. O ponto é que isso também traz questionamentos sobre quando e como dar a devida transparência no uso da GenAI. O próximo capítulo trata dessas questões.

4.2 Diretrizes Práticas para a Declaração de Uso de IA

Tanto em produções acadêmicas quanto em materiais didáticos, ainda não existe uma norma universalmente aceita que oriente a forma adequada de declarar o uso de sistemas de inteligência artificial generativa. Entretanto, para promover a auditabilidade, a responsabilidade autoral e a transparência metodológica, é fundamental que tais declarações sejam elaboradas de modo preciso, específico e verificável.

Com esse propósito, foram desenvolvidos alguns modelos que buscam padronizar e facilitar o processo de *disclosure* do uso de IA em contextos científicos e educacionais, promovendo maior transparência, ética e confiabilidade na produção acadêmica. Esses modelos têm como objetivo orientar pesquisadores, docentes e instituições na forma adequada de declarar a utilização de ferramentas de IA, garantindo que o leitor compreenda claramente em que etapas e com quais finalidades tais recursos foram empregados.

Nesse contexto, entre os modelos mais amplamente reconhecidos destaca-se o AID Framework, uma abreviação de

Artificial Intelligence Disclosure Framework, que se consolidou como uma referência internacional por oferecer diretrizes práticas e estruturadas para a comunicação responsável do uso de IA. O próximo capítulo apresenta esse modelo com mais detalhes, explorando seus princípios, aplicações e benefícios para a comunidade científica e educacional.

4.2.1 AID framework

Este modelo foi proposto por Kari D. Weaver da Universidade de Waterloo em 2024, como uma alternativa aos modelos tradicionais de citação. O seu objetivo central é aumentar a transparência, a rastreabilidade e a responsabilidade no uso de GenAI em atividades de pesquisa, escrita e ensino (Weaver, 2024).

O modelo surge da constatação de que as normas tradicionais de citação não são suficientes para descrever adequadamente a interação dinâmica entre autor e IA, já que os resultados gerados podem variar conforme o modelo, o prompt e os parâmetros utilizados. Com isso, o AID Framework adapta o conceito de “papéis de contribuição” — originalmente usado para descrever a função de coautores — para caracterizar o envolvimento de sistemas de IA nos diferentes estágios da produção intelectual.

Para isso, esse framework estabelece que a declaração de uso de IA (AID Statement) seja inserida ao final do trabalho, de forma semelhante a uma seção de agradecimentos. Essa declaração deve ser

concisa, específica e legível tanto para humanos quanto para máquinas, seguindo uma estrutura padronizada composta por pares de cabeçalho e descrição.

A estrutura geral de uma declaração de uso de IA, portanto, deve seguir o padrão:

Declaração de AID:

Ferramenta de Inteligência Artificial: [descrição das ferramentas utilizadas];

[Título]: [descrição do uso da IA nessa etapa do trabalho];

[Título]: [descrição do uso da IA nessa etapa do trabalho].

...

Cada par “Título: Descrição” deve terminar com ponto e vírgula, exceto o último, que termina com ponto final (Weaver, 2024). O **Título** se refere à categoria aplicável, cujas opções possíveis estão entre as 14 categorias que cobrem todas as etapas do ciclo de produção acadêmica. O autor deve incluir em sua declaração apenas aquelas que se aplicam ao seu caso. Segundo Weaver (2024), essas categorias são:

1. **Recurso(s) de IA:** para essa categoria, o autor deve descrever nome, versão e data de uso do recurso usado. Podem ser descritas limitações ou vieses conhecidos.

2. **Conceituação:** deve ser declarado caso o autor tenha feito uso de tecnologias de IA para elaboração ou definição do tema, problema, objetivos, justificativa, hipóteses ou outras etapas preliminares de planejamento da pesquisa.

Exemplos:

- *Usar IA para sugerir hipóteses sobre o impacto da urbanização na biodiversidade.*
- *Pedir à IA para reformular perguntas de pesquisa para torná-las mais específicas ou mensuráveis.*
- *Pedir à IA para formular o objetivo da pesquisa com base em um dado problema estabelecido.*

3. **Metodologia:** deve ser declarado caso o autor tenha aplicado tecnologias de IA para apoio na definição do desenho da pesquisa. Sugestões de abordagem ou procedimentos metodológicos, critérios de seleção de amostras etc.

Exemplos:

- *IA sugere o uso de estudo de caso em vez de experimento controlado.*
- *IA ajuda a definir critérios de inclusão/exclusão de participantes.*
- *Pesquisador solicita ajuda à IA para determinar o tamanho da amostra de participantes, com base em uma determinada população.*

4. **Coleta de Informações:** deve ser declarado quando há uso de IA para identificar padrões na literatura existente e reunir informações relevantes para a formulação, desenvolvimento ou desenho do estudo.

Exemplos:

- *IA analisa centenas de artigos para identificar lacunas sobre saúde mental em adolescentes.*
- *IA sugere autores-chave e teorias relevantes para um dado tema.*

5. **Métodos de Coletas de Dados:** deve ser declarado quando há uso de IA no desenvolvimento de ferramentas ou instrumentos usados para coletar dados para o estudo.

Exemplos:

- *Criar um chatbot com IA para aplicar questionários.*
- *Desenvolver um app com IA para registrar dados de sono via sensores.*

6. **Execução:** deve ser declarado quando há aplicações diretas de IA em procedimentos de pesquisa.

Exemplos:

- *IA coleta dados automaticamente de sites públicos sobre preços de alimentos.*
- *Aplicar uma pesquisa gerada e distribuída por IA para simular respostas de consumidores.*

7. **Curadoria de Dados:** deve ser declarado quando é feito uso de IA em processos que envolvem organizar, limpar, manter e tornar os dados acessíveis e úteis para análise ou reutilização.

Exemplo:

- *Aplicar IA para identificar valores duplicados, erros de digitação e inconsistências em conjuntos de dados oriundos de formulários de pesquisa.*

8. **Análise de Dados:** deve ser declarado quando há uso de IA para a realização de análises estatísticas ou matemáticas, regressões, análise de texto etc.

Exemplos:

- *Aplicar aprendizado de máquina para prever desempenho escolar de estudantes.*
- *Usar IA para classificar sentimentos em comentários de redes sociais.*

9. **Privacidade e Segurança:** deve-se declarar sempre que forem utilizados métodos de IA para garantir a proteção da privacidade e a segurança dos dados, assegurando que tais práticas estejam em conformidade com as diretrizes legais, disciplinares, éticas e políticas institucionais.

Exemplo:

- *Usar IA para anonimizar dados de participantes antes de uma análise.*

10. **Interpretação:** deve-se declarar, sempre que for feito uso de IA para categorizar, resumir ou manipular dados e sugerir conclusões associadas.

Exemplos:

- *Aplicar IA para resumir os principais achados de uma análise estatística.*
- *Usar IA para sugerir implicações políticas e sociais com base nos dados coletados.*

11. **Visualização:** deve ser declarado quando forem aplicadas tecnologias de IA para a criação de gráficos e representações visuais.

12. **Escrita, Revisão e Edição:** deve ser declarado quando forem aplicadas tecnologias de IA para fazer revisões, sugestões estilísticas ou estruturais no texto.

Exemplos:

- *Correção de estilo acadêmico.*
- *Melhora da coesão textual, sugestões de cortes de trechos redundantes etc.*

13. **Tradução:** deve ser declarado quando houver processos de tradução realizada ou auxiliada por tecnologias de IA.

14. **Administração do Projeto:** deve ser declarado quando forem executadas quaisquer tarefas administrativas relacionadas ao estudo, incluindo o gerenciamento de orçamentos, cronogramas e comunicações.

Veja a seguir um exemplo de aplicação de declaração de uso de IA usando esse framework no contexto da pesquisa.

Declaração de Uso de IA:

Ferramenta de IA: ChatGPT (GPT-4o, OpenAI, outubro de 2025);

Conceitualização: ChatGPT foi utilizado para revisar questões de pesquisa;

Análise de Dados: ChatGPT foi empregado para verificar temas codificados em respostas abertas;

Escrita, Revisão e Edição: ChatGPT auxiliou na revisão de sentenças e metáforas na seção de revisão de literatura.

É preciso que haja um conjunto “título:descrição” para cada ferramenta utilizada durante o processo (Weaver, 2024). Veja a seguir um exemplo de declaração para uma atividade educacional.

Declaração de Uso de IA:

Ferramenta de IA: Microsoft Copilot;

Conceitualização: o Copilot foi usado para elaborar o plano de aula.

Coleta de Informações: o Copilot foi usado para localizar artigos científicos;

Visualização: o Copilot foi usado para gerar gráficos comparativos dos resultados obtidos.

Dada a sua capacidade de descrever de forma estruturada e transparente todo o fluxo de um processo de pesquisa, esse framework mostra-se particularmente útil na declaração de uso da IA. Ele permite registrar de maneira clara em quais etapas a IA foi aplicada, desde a coleta e curadoria de dados até a análise, interpretação e escrita dos resultados, garantindo maior rastreabilidade e credibilidade científica. Além disso, contribui para a padronização das práticas de *disclosure*, fortalecendo a ética acadêmica e oferecendo aos leitores e avaliadores uma visão abrangente sobre como os recursos tecnológicos influenciaram o desenvolvimento da produção.

4.2.2 *Simplificando as coisas*

Ainda buscando compreender como e quando a GenAI deve ser declarada em trabalhos acadêmicos e artigos científicos, os pesquisadores David B. Resnik e Mohammad Hosseini (2025), propuseram critérios práticos e normativos para orientar pesquisadores, revisores e editores sobre estas questões. Para isso, realizaram uma análise comparativa de políticas editoriais de grandes grupos e instituições científicas, como *Nature*, *Springer*, *Taylor & Francis*, *World Association of Medical Editors (WAME)* e *International Committee of Medical Journal Editors (ICMJE)*.

A partir dessa análise, Resnik e Hosseini (2025) identificam divergências significativas nas exigências de disclosure, que variam

desde a obrigatoriedade de relatar qualquer uso de IA até a dispensa de menção quando o uso é considerado trivial.

Para preencher essa lacuna de uniformidade, os autores formulam um modelo de decisão baseado em dois critérios principais: intencionalidade e substancialidade. O primeiro refere-se ao uso deliberado da IA com uma finalidade específica no processo de pesquisa ou redação; o segundo avalia o impacto real desse uso sobre o conteúdo, os resultados ou as conclusões do trabalho.

Com base nesses critérios, foram propostas três categorias de necessidade para a declaração de uso de IA: obrigatória, opcional e desnecessária. Veja a seguir, mais detalhes sobre cada uma delas (Hosseini; Resnik 2025):

1. **Declaração obrigatória:** deve ocorrer quando o uso da IA tem um impacto substancial no conteúdo científico ou nos resultados obtidos. Isso inclui situações em que a IA é utilizada para gerar textos, imagens, dados sintéticos ou tabelas, bem como quando desempenha papéis analíticos, como classificar dados, realizar mineração textual ou interpretar imagens. Em suma, a declaração é obrigatória quando a GenAI for usada para:
 - a. *Formulação de hipóteses ou desenho de experimentos:* se a IA ajudou a definir perguntas de pesquisa ou planejar experimentos, isso deve ser declarado.

- b. *Redação ou revisão substancial de texto*: quando a IA é usada para redigir partes do artigo, resumir, parafrasear ou integrar conteúdo de forma significativa.
- c. *Tradução de partes ou do artigo inteiro*: se a IA foi usada para traduzir o texto científico, isso é considerado uma contribuição substancial.
- d. *Análise ou visualização de dados*: inclui o uso de IA para interpretar dados quantitativos ou qualitativos ou para gerar gráficos e tabelas que sustentam os resultados.
- e. *Extração de dados para revisão de literatura*: por exemplo, usar IA para selecionar artigos relevantes ou identificar lacunas no conhecimento.
- f. *Geração de dados sintéticos ou imagens*: se a IA criou dados ou imagens que foram incluídos na pesquisa ou publicação.

Nesses casos, a omissão seria eticamente problemática, pois a IA passa a integrar o processo criativo e decisório que fundamenta a validade científica do estudo. Além de comprometer a transparência, tal ausência pode gerar dúvidas quanto à originalidade dos resultados, dificultar a reprodutibilidade das análises e enfraquecer a confiança da comunidade acadêmica. Declarar o uso da IA, portanto, não é apenas uma questão de conformidade, mas um requisito essencial para assegurar integridade, credibilidade e responsabilidade na produção científica.

2. **Declaração opcional:** abrange usos auxiliares que não afetam de maneira significativa o conteúdo ou as conclusões do trabalho. Por exemplo:

- a. *Revisão de texto para gramática, ortografia ou organização:* quando a IA é usada apenas para corrigir erros ou melhorar a estrutura de frases escritas por humanos.
- b. *Busca e verificação de referências:* utilizar IA para encontrar fontes ou confirmar a relevância de referências já encontradas por humanos.
- c. *Geração de exemplos para conteúdo existente:* por exemplo, pedir à IA que sugira exemplos ilustrativos para conceitos já desenvolvidos no texto.
- d. *Sugestões para organização ou títulos:* usar IA para brainstorm de estrutura do artigo ou ideias de títulos para seções.
- e. *Validação ou feedback sobre ideias, textos ou códigos:* quando a IA é usada como apoio para revisar ou comentar sobre conteúdo já criado por humanos.

Nesses casos, embora não exista uma obrigação ética formal de divulgar o uso, os autores consideram recomendável fazê-lo, pois a transparência reforça a confiança no processo editorial e reduz a possibilidade de interpretações equivocadas sobre a autoria do texto.

3. **Declaração desnecessária:** aplica-se a situações em que o uso da IA é trivial ou inerente a softwares amplamente utilizados. Isso inclui, por exemplo:
- a. *Sugestão de palavras ou frases para melhorar clareza: quando a IA apenas propõe ajustes pontuais em frases já escritas por humanos, sem alterar o conteúdo ou significado.*
 - b. *Uso incidental em sistemas integrados: por exemplo, quando a IA está embutida em ferramentas como buscadores ou equipamentos laboratoriais (como sequenciadores genômicos), mas não é usada de forma direta ou intencional pelo pesquisador.*
 - c. *Assistência digital para organização de arquivos e fluxos de trabalho: utilizar IA como apoio administrativo — por exemplo, para gerenciar pastas, backups ou cronogramas — sem impacto no conteúdo científico.*

Esses casos são considerados triviais ou administrativos, e não influenciam o conteúdo intelectual da pesquisa, a exigência de declaração seria excessiva e sem relevância ética. Além disso, incluí-los no *disclosure* pode dificultar que o leitor ou avaliador possa focar nos pontos realmente relevantes do trabalho.

Para facilitar a compreensão sobre os momentos adequados para declarar ou não o uso de IA, propõe-se uma síntese ilustrativa. A Figura 1 apresenta essa síntese, indicando de forma clara como proceder.

Figura 1. Quando declarar uso de IA.



Fonte: Produzido pelos autores, baseado em (Hosseini; Resnik 2025).

Quanto ao que deve ser dito no *disclosure*, é proposta também uma forma mais simplificada e direta de descrever o uso de tecnologias de IA, sem a necessidade de estabelecer critérios excessivamente rigorosos ou explicitar detalhadamente as fases do projeto como ocorre no AID Framework.

Nesse modelo alternativo, é indicado que sejam informados apenas o nome da ferramenta (e o desenvolvedor, quando aplicável), a versão utilizada ou a data de acesso, a natureza específica da contribuição da IA, bem como a garantia de que houve revisão e validação humana do conteúdo gerado. Veja exemplos a seguir:

Exemplo 1: “Partes da introdução foram geradas com o auxílio do ChatGPT (OpenAI, versão 4, acesso em março de 2025), revisadas e editadas pelos autores para precisão, clareza e adequação científica”.

Exemplo 2: “A análise de dados foi realizada com suporte de um modelo de aprendizado de máquina (BERT), utilizado para a classificação automática dos documentos. Todos os resultados foram posteriormente verificados e validados pelos pesquisadores”.

Exemplo 3: “As imagens da Figura 2 foram criadas com auxílio do DALL-E (OpenAI, versão 3), com base em descrições elaboradas pelos autores. As ilustrações foram revisadas pelos pesquisadores quanto à fidelidade conceitual e ausência de distorções”.

Em síntese, a prática de declarar o uso de GenAI em trabalhos acadêmicos deve ser orientada pela ética, pela transparência e pela rastreabilidade, permitindo que o leitor compreenda o papel efetivo da tecnologia na construção do conhecimento apresentado.

4.2.3 O que não fazer

Os modelos de declaração de uso de IA tratados nas seções anteriores dão uma ideia da complexidade e falta de padronização e consenso no que diz respeito à realização deste trabalho. Por isso, é importante sempre estar atento às normas da instituição, editora ou periódico para o qual se pretenda produzir algum trabalho. Este livro

apresenta alguns caminhos possíveis, mas estes caminhos não substituem as exigências específicas de cada instituição.

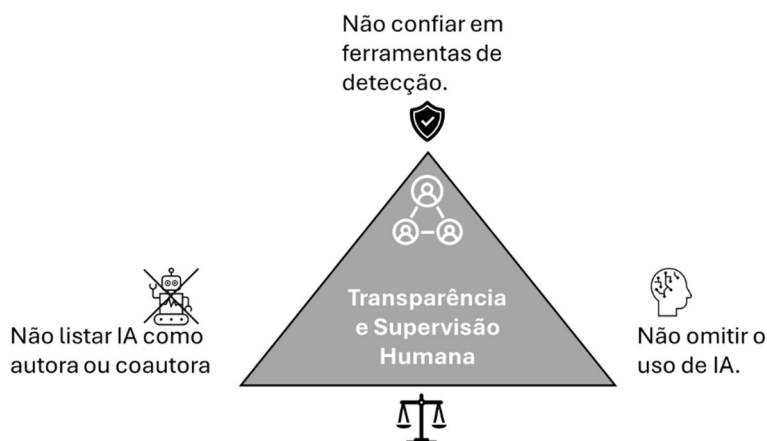
Além disso, o imperativo ético exige que se adote uma postura de honestidade intelectual, transparência e bom senso, o que implica na rejeição de práticas como:

- 4. Listar a IA como autor ou coautor:** a Inteligência Artificial deve ser encarada estritamente como uma ferramenta de apoio, e não como um agente responsável pela criação ou concepção intelectual do trabalho. A responsabilidade e a validação do conteúdo devem sempre recair sobre o(s) autor(es) humano(s), sendo incorreto ou inapropriado atribuir autoria a um software. Observar este critério é mandatório, pois muitas editoras, conforme já mencionado, trazem normas explícitas sobre a não aceitação dessa prática.
- 5. Omitir a utilização de IA:** tentar esconder o uso de ferramentas de IA para a geração, estruturação ou aprimoramento do texto, na esperança de evitar a detecção, viola diretamente os princípios de integridade acadêmica e de honestidade. A divulgação transparente é obrigatória, tanto quanto a descrição adequada da metodologia, para que o leitor e os avaliadores compreendam o processo de produção do trabalho.

6. Confiar em ferramentas de detecção de IA: não se deve confiar exclusivamente em softwares projetados para identificar conteúdo gerado por IA. Essas ferramentas frequentemente apresentam uma alta taxa de erro (falsos positivos), confundindo, muitas vezes, textos humanos originais com textos artificiais, o que pode levar a avaliações e penalidades injustas. A verificação deve ser feita por meio de julgamento humano e análise crítica (Weber-Wulff *et al.*, 2023; Dalalah; Dalalah, 2023; Elkhatat *et al.*, 2023).

A Figura 2, apresentada a seguir, sintetiza o que não deve ser feito ao utilizar tecnologias de inteligência artificial na produção de conhecimento ou na criação de conteúdo sintético.

Figura 2: O que não fazer ao usar IA na produção de conteúdo.



Fonte: Produzido pelos autores.

Todas essas ações visam mitigar o mau uso da GenAI que pode resultar em plágio, desinformação ou discriminação, muitas vezes sem que o autor esteja consciente. A emergência destas novas tecnologias impõe um reexame fundamental das práticas de integridade acadêmica e profissional.

O princípio central da responsabilidade intelectual é inegociável: a IA é uma ferramenta poderosa, mas inerentemente falível (propensa a alucinações e vieses), o que torna a agência humana — o pensamento crítico e a verificação factual — o único fator capaz de garantir a confiabilidade, a precisão e a ética na produção e disseminação de conhecimento.

GLOSSÁRIO

1. **Agente (de IA):** sistema de software ou hardware (como um chatbot, tutor inteligente ou robô) capaz de perceber seu ambiente, tomar decisões com base em suas percepções e agir para alcançar seus objetivos com relativa autonomia.
2. **Algoritmo:** conjunto finito de regras e instruções sequenciais necessários para atingir um objetivo. No contexto da IA, é um conjunto de instruções escritas em linguagem de programação para definir operações que o computador deve seguir. Pode implementar modelos matemáticos e regras complexas.
3. **Alucinação (da IA):** geração de informações factualmente falsas, referências inexistentes ou dados inventados por modelos de GenAI, comprometendo a confiabilidade dos resultados gerados.
4. **Datasets:** grandes conjuntos de dados (textos, imagens, códigos, obras) usados para treinar os modelos de IA.
5. **Direitos Autorais:** conjunto de proteções legais que garante ao criador o controle sobre o uso e a reprodução de sua obra. A lei brasileira exige Autoria Humana para sua concessão.
6. **Governança em IA:** conjunto de estruturas, práticas e princípios destinados a orientar o desenvolvimento, uso e supervisão da IA de forma responsável, ética e transparente. Deve envolver todas as partes interessadas.
7. **Grande Modelo de Linguagem (LLM):** modelos de IA treinados em vastos conjuntos de dados para gerar e processar linguagem natural. São a base da GenAI.
8. **GPT (Generative Pre-trained Transformer):** arquitetura de modelo de Machine Learning baseada em Transformers,

amplamente utilizada para IA Generativa, sendo o motor por trás de ferramentas como o ChatGPT.

9. **IA Generativa (GenAI / IAGen):** subárea da IA focada no desenvolvimento de modelos capazes de gerar conteúdo sintético (texto, código, imagem).
10. **Inteligência Artificial (IA):** área da Ciência da Computação dedicada a desenvolver técnicas para possibilitar que máquinas possam imitar artificialmente atributos da inteligência humana como aprendizagem, raciocínio, inferência etc. Visa desenvolver modelos computacionais capazes de imitar um ser humano em tarefas cognitivas, como ler e interpretar linguagens, aprender, decidir logicamente, entre outras.
11. **Machine Learning (Aprendizado de Máquina):** subárea da IA que busca desenvolver técnicas que permitam simular artificialmente o aprendizado em máquinas. Possibilita que sistemas possam melhorar seus resultados a partir da experiência (dados) sem serem explicitamente programados.
12. **Modelo (de IA):** resultado de um processo de aprendizado de máquina sobre um conjunto de dados (dataset). Um modelo representa os padrões e relações aprendidos a partir dos dados de treinamento e é usado para fazer previsões, classificações, ou gerar novos dados
13. **Propriedade Intelectual (PI):** direitos legais que protegem criações do intelecto humano (textos, obras, designs).
14. **Vieses de IA:** tendências da IA em repetir erros ou desigualdades presentes no conjunto de dados usado para seu treinamento. Efeito colateral do aprendizado de máquinas.

DISCLOSURE (DECLARAÇÃO DE USO DE IA)

Ferramenta de IA: Scyspace, plataforma de apoio à pesquisa científica com recursos de IA, da Scispace. Modelo não especificado. Novembro de 2025.

Conceituação: o Scyspace foi utilizado para explorar diferentes fontes acadêmicas sobre os temas tratados no livro.

Ferramenta de IA: Gemini, assistente de IA Generativa da Google, modelo 2.5 Flash. Novembro de 2025.

Conceitualização: o Gemini auxiliou no processo de busca profunda sobre tópicos relacionados ao tema do livro e fornecendo ideias para a estrutura de capítulos.

Coleta de Informações: o Gemini auxiliou na identificação de padrões em artigos científicos e relatórios técnicos, reunindo informações e conceitos relevantes que embasaram o desenvolvimento de capítulos específicos e fortaleceram o arcabouço teórico da obra.

Escrita, Revisão e Edição: o Gemini auxiliou na revisão de alguns parágrafos em diferentes capítulos, na busca por maior coesão e clareza textual.

Todo o conteúdo foi escrito, revisado e aprovado pelos autores.

Ferramenta de IA: ChatGPT, da OpenAI, modelo GPT-5.2. Dezembro de 2025).

Escrita, Revisão e Edição: ChatGPT auxiliou na revisão de alguns parágrafos em diferentes capítulos, na busca por maior coesão e clareza textual.

Todo o conteúdo foi escrito, revisado e aprovado pelos autores.

Ferramenta de IA: ChatGPT, modelo de geração de imagens da OpenAI (*gpt-image-1*). Dezembro de 2025.

Visualização: ChatGPT auxiliou na criação conceitual da figura na capa do livro.

REFERÊNCIAS

AHMAD, Sayed Fayaz et al. Impact of artificial intelligence on human loss in decision making, laziness and safety in education. **Humanities and Social Sciences Communications**, v. 10, n. 311, 2023.

Disponível em: < <https://doi.org/10.1057/s41599-023-01787-8> >.

Acesso em: 01 dez. 2025.

ANANYA. AI image generators often give racist and sexist results: can they be fixed? **Nature**, Londres, v. 627, p. 722–725, 19 mar. 2024.

Disponível em: <<https://www.nature.com/articles/d41586-024-00674-9>>. Acesso em: 21 out. 2025.

ARDITO, Cesare G. Contra generative AI detection in higher education assessments. **arXiv preprint arXiv:2312.05241**, 2023. Disponível em: <<https://arxiv.org/abs/2312.05241>>. Acesso em: 20 out. 2025.

ATARI, Mohammad; XUE, Mona J.; PARK, Peter S.; BLASI, Damián; HENRICH, Joseph. Which Humans? **PsyArXiv Preprints**, Charlottesville: Center for Open Science, 2023. Disponível em: <https://osf.io/preprints/psyarxiv/5b26t_v1>. Acesso em: 1 dez. 2025.

AUSTRALIAN DEPARTMENT OF EDUCATION. **Australian Framework for Generative Artificial Intelligence in Schools**. Australia: Commonwealth of Australia, 2023. Disponível em: <<https://www.education.gov.au/schooling/resources/australian-framework-generative-artificial-intelligence-ai-schools>>. Acesso em: 14 out. 2025.

BANCO MUNDIAL. **AI revolution in education: what you need to know**. Washington, DC: World Bank Publications, 2024. Acesso em: < [Artificial Intelligence Revolution in Education: What You Need to Know](#) >

BRASIL. Presidência da República. **Lei nº 8.069, de 13 de julho de 1990. Dispõe sobre o Estatuto da Criança e do Adolescente e dá outras providências**. Diário Oficial da União: seção 1, Brasília, DF, 16 jul. 1990. Disponível em:

<https://www.planalto.gov.br/ccivil_03/leis/l8069.htm>. Acesso em: 16 out. 2025.

BRASIL. Presidência da República. **Lei nº 13.709, de 14 de agosto de 2018. Dispõe sobre a proteção de dados pessoais e altera a Lei nº 12.965, de 23 de abril de 2014 (Marco Civil da Internet).** Diário Oficial da União: seção 1, Brasília, DF, 15 ago. 2018. Disponível em: <https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm>. Acesso em: 16 out. 2025.

CIEB. Notas Técnicas **#21: Inteligência artificial na educação básica: novas aplicações e tendências para o futuro.** São Paulo, SP: CIEB, 2024.

CÓRDOVA, Paulo Roberto. **Ethoscool: avançando a ética em IA na educação por meio de agentes morais artificiais.** 2023. 136 f. Tese (Doutorado em Informática na Educação) — Programa de Pós-Graduação em Informática na Educação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2023.

CÓRDOVA, Paulo Roberto. **Inteligência Artificial: entre o fascínio e o medo.** São Paulo: Editora Contexto, 2025.

DAHMANI, Louisa; BOHBOT, Véronique D. Habitual use of GPS negatively impacts spatial memory during self-guided navigation. **Scientific Reports**, Londres: Springer Nature, v. 10, n. 6310, 2020. Disponível em: < <https://doi.org/10.1038/s41598-020-62877-0>>. Acesso em: 1 dez. 2025.

Dalalah, Doraid; Dalalah, Osama M. A. The false positives and false negatives of generative AI detection tools in education and academic research: the case of ChatGPT. **The International Journal of Management Education**, [S. l.], v. 21, 100822, 2023. Disponível em: < <https://doi.org/10.1016/j.ijme.2023.100822> >. Acesso em: 02 dez. 2025.

ELKHATAT, Ahmed M.; ELSAID, Khaled; ALMEER, Saeed. Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text. **International Journal for Educational Integrity**, [S. l.], v. 19, n. 17, 2023. Disponível em: < <http://dx.doi.org/10.1007/s40979-023-00140-5>>. Acesso em: 02 dez. 2025.

ELSEVIER. **Generative AI policies for journals**. Amsterdã: Elsevier, 2025. Disponível em: <<https://www.elsevier.com/about/policies-and-standards/generative-ai-policies-for-journals>>. Acesso em: 1 dez. 2025.

GONG, Chen; YANG, Yang. Google effects on memory: a meta-analytical review of the media effects of intensive Internet search behavior. **Frontiers in Public Health**, [S. l.], v. 12, p. 1332030, 18 jan. 2024. Disponível em: <https://doi.org/10.3389/fpubh.2024.1332030>. Acesso em: 20 out. 2025.

HAIDER, Jutta; SÖDERSTRÖM, Kristofer Rolf; EKSTRÖM, Björn; RÖDL, Malte. GPT-fabricated scientific papers on Google Scholar: key features, spread, and implications for preempting evidence manipulation. **Harvard Kennedy School (HKS) Misinformation Review**, 2024. Disponível em: < <https://doi.org/10.37016/mr-2020-156>>. Acesso em: 28 out. 2025.

HORODETSKA, Natella. Mental arithmetic and its impact on the development of mathematical abilities. **The American Journal of Social Science and Education Innovations**, [S. l.], v. 7, n. 4, p. 48–54, 27 abr. 2025. Disponível em: < <https://doi.org/10.37547/tajssei/Volume07Issue04-07>>. Acesso em: 20 out. 2025.

IEEE – Institute of Electrical and Electronics Engineers. **The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems**. Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems. 1. ed. 2019. Disponível em: <<https://standards.ieee.org/content/ieee-standards/en/industry-connections/ec/autonomous-systems.htm>>. Acesso em: 05 set. 2025.

KOSMYNA, Nataliya et al. Your brain on ChatGPT: accumulation of cognitive debt when using an AI assistant for essay writing task. **arXiv preprint arXiv:2506.08872**, 2025. Disponível em: < <https://doi.org/10.48550/arXiv.2506.08872>>. Acesso em: 22 set. 2025.

LEMANN. FUNDAÇÃO LEMANN; TELLES FOUNDATION. **Perspectivas sobre IA e Educação Básica**: Aliança IA para Educação. São Paulo: Fundação Lemann, 2024. Disponível em: <<https://iaparaeducacao.org.br/>>. Acesso em: 20 out. 2025.

LIN, Hsiu-Man et al. Structural and functional neural correlates in individuals with excessive smartphone use: a systematic review and meta-analysis. **International Journal of Environmental Research and Public Health**, Basel, v. 19, n. 23, p. 16277, 5 dez. 2022. Disponível em: <<https://doi.org/10.3390/ijerph192316277>>. Acesso em: 20 out. 2025.

MAURYA, Rajesh Kumar; MAURYA, Swati R. Content reliability in the age of AI: a comparative study of human vs. GPT-generated scholarly articles. **Library Progress International**, v. 44, n. 3, p. 1932–1943, jul./dez. 2024. Disponível em: <<https://doi.org/10.48165/bapas.2024.44.2.1>>. Acesso em: 02 dez. 2025.

MUSA, Nazaruddin; ISHAK, Mohd Sobhi. The phenomenon of Google effect, digital amnesia and nomophobia in academic perspective. **Cyberspace: Jurnal Pendidikan Teknologi Informatika**, [S. l.], v. 5, n. 1, p. 1–15, mar. 2021. Acesso em: <<https://doi.org/10.22373/cj.v5i1.8219>>. Acesso em: 21 out. 2025.

NEES. **Inteligência Artificial Generativa na Educação**. Maceió: Núcleo de Excelência em Tecnologias Sociais – NEES, 2024. Disponível em: <<https://iaedu.nees.ufal.br/wp-content/uploads/2025/04/NT-1-Inteligencia-Artificial-Generativa-na-Educacao.pdf>>. Acesso em: 02 dez. 2025.

OCDE. **The potential impact of artificial intelligence on education: opportunities and challenges**. Paris: OCDE Publishing, 2024. Disponível em: <https://www.oecd.org/en/publications/the-potential-impact-of-artificial-intelligence-on-equity-and-inclusion-in-education_15df715b-en.html>. Acesso em: 17 out. 2024.

RESNIK, David B.; HOSSEINI, Mohammad. Disclosing artificial intelligence use in scientific research and publication: When should disclosure be mandatory, optional, or unnecessary? **Accountability in Research**, [S. l.], v. 32, n. 3, p. 1–25, 2025. Disponível em: <<https://doi.org/10.1080/08989621.2025.2481949>>. Acesso em: 22 out. 2025.

SABA, Walid. LLMs' Understanding of Natural Language Revealed. **arXiv preprint arXiv:2407.19630**. Ithaca: Cornell University, 2024. Disponível em: <<https://doi.org/10.48550/arXiv.2407.19630>>. Acesso em: 1 dez. 2025.

Salles, A., Evers, K., & Farisco, M. (2020). Anthropomorphism in AI. **AJOB Neuroscience**, 11(2), 88–95. Disponível em: <<https://doi.org/10.1080/21507740.2020.1740350>>. Acesso em: 15 out. 2025.

SPRINGER NATURE LIMITED. **Artificial Intelligence (AI)**. Londres: Springer Nature, 2025. Disponível em: <<https://www.nature.com/nature-portfolio/editorial-policies/ai>>. Acesso em: 1 dez. 2025.

TAYLOR & FRANCIS. **AI Policy**. Londres: Taylor & Francis, 2025. Disponível em: <<https://taylorandfrancis.com/our-policies/ai-policy>>. Acesso em: 1 dez. 2025.

UNESCO. **AI competency framework for students**. Paris: UNESCO, 2024. DOI: 10.54675/JKJB9835. Disponível em: <<https://unesdoc.unesco.org/ark:/48223/pf0000391105>>. Acesso em: 19 out. 2025.

UNESCO. **Generative AI and the future of education**. Paris: UNESCO, 2023. Disponível em: <<https://unesdoc.unesco.org/ark:/48223/pf0000385877>>. Acesso em: 14 set. 2025.

UNESCO. **Recommendation on the ethics of artificial intelligence**. Paris: UNESCO, 2022. Disponível em: <<https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence>> Acesso em: 02 dez. 2025.

U.S. DEPARTMENT OF EDUCATION. **Artificial Intelligence and the future of teaching and learning**: insights and recommendations. Washington, DC: Department of Education, 2023. Disponível em: <<https://www.ed.gov/sites/ed/files/documents/ai-report/ai-report.pdf>>. Acesso em: 02 dez. 2025.

VICARI, Rosa Maria; BRACKMANN, Christian; MIZUSAKI, Lucas; GALAFASSI, Cristiano. **Inteligência Artificial na Educação Básica**: prática na escola. São Paulo, SP: Novatec, 2023.

Watson, D. The Rhetoric and Reality of Anthropomorphism in Artificial Intelligence. **Minds & Machines** **29**, 417–440 (2019). Disponível em: <https://doi.org/10.1007/s11023-019-09506-6>>. Acesso em: 28 out. 2025.

WEBER WULFF, Debora et al. Testing of detection tools for AI generated text. **International Journal for Educational Integrity**, [S. l.], v. 19, n. 26, 2023. Disponível em: <<https://doi.org/10.1007/s40979-023-00146-z>>. Acesso em: 02 dez. 2025.

WEAVER, Kari D. The Artificial Intelligence Disclosure (AID) Framework: an introduction. **College & Research Libraries News**, v. 85, n. 10, 2024. Disponível em: <<https://doi.org/10.5860/crln.85.10.407>>. Acesso em: 22 out. 2025.

Sobre os Autores



Dr. Paulo Roberto Córdova é professor do Instituto Federal de Educação, Ciência e Tecnologia de Santa Catarina (IFSC). Doutor em Informática na Educação pela Universidade Federal do Rio Grande do Sul - UFRGS, mestre em Desenvolvimento e Sociedade e bacharel em Sistemas de Informação. Sua atuação se concentra em Engenharia de Software e Inteligência Artificial Aplicada, com foco em ética, governança e nos impactos das tecnologias emergentes na sociedade e, de modo particular, na educação.



Dra. Rosa Maria Vicari é professora titular do Instituto de Informática da Universidade Federal do Rio Grande do Sul (UFRGS). Doutora em Engenharia Eletrotécnica e de Computadores pela Universidade de Coimbra, mestre em Ciência da Computação e graduada em Processamento de Dados. Sua atuação concentra-se em Inteligência Artificial aplicada à educação, especialmente em sistemas tutores inteligentes e ambientes de aprendizagem mediada por tecnologia.

QUEM ESCREVE O CONHECIMENTO?

A Inteligência Artificial não está apenas transformando a educação — está redesenhando seus fundamentos. Em um cenário em que algoritmos escrevem, avaliam e criam ao nosso lado, a fronteira entre autoria humana e produção automatizada se torna cada vez mais tênue. É justamente nesse território instável que este livro se posiciona.

Com olhar crítico e profundo, os autores examinam as tensões que emergem quando ferramentas generativas entram na sala de aula, nos laboratórios de pesquisa e no próprio ato de aprender. Aqui, discutem-se os dilemas éticos que acompanham essa revolução: a erosão da autoria original, os vieses que moldam decisões automatizadas, o impacto cognitivo de delegarmos habilidades essenciais e a urgente necessidade de transparência no uso da IA. Mais do que um diagnóstico, este livro oferece caminhos. Propõe princípios de governança, critérios de uso responsável e orientações práticas para integrar a IA sem comprometer aquilo que define a essência da educação: o pensamento crítico, a autonomia intelectual e o desenvolvimento humano.

Trata-se de um convite — e um desafio. Ler estas páginas é entrar em um debate que ultrapassa modismos tecnológicos. É decidir que tipo de futuro educacional queremos construir antes que os algoritmos o façam por nós.

