DEVIR EDUCAÇÃO
ISSN: 2526-849X

Vieses, ilusões e decisões: até que ponto podemos confiar na inteligência artificial?

Biases, illusions, and decisions: to what extent can we trust artificial intelligence?

Sesgos, ilusiones y decisiones: ¿hasta qué punto podemos confiar en la inteligencia artificial?

Juliana Pires da Silva[1]
Priscila Flores da Luz[2]
Rafael Arlindo Rosa[3]

Resumo

A Inteligência Artificial (IA), desenvolvida desde a década de 1950, é uma tecnologia de propósito geral que requer análise crítica de suas implicações éticas, epistemológicas e sociotécnicas. Embora prometa inovação e eficiência, suscita preocupações como vieses algorítmicos, bolhas sociais, alucinações e opacidade dos modelos. Neste ensaio teórico, a confiabilidade da IA é discutida a partir de três eixos: os vieses, as "ilusões" e as decisões tomadas com base em suas respostas. Para isso, adota-se uma abordagem metodológica qualitativa, centrada na revisão de literatura, dialogando com autores como Rivoltella (2024) e Fantin (2024). Mostra-se que os vieses algorítmicos refletem preconceitos presentes nos dados de treinamento, como ocorre com o ChatGPT e sistemas de reconhecimento facial que falham na identificação de rostos negros, reforçando culturas hegemônicas e afetando o campo educacional. As "ilusões" contribuem para a disseminação de desinformação e fortalecem a pós-verdade. Neste cenário, a literacia em IA torna-se essencial para que os usuários reconheçam limitações e falhas. As decisões automatizadas baseiam-se frequentemente em modelos opacos, que comprometem a explicabilidade e demandam mediação humana ativa. Ainda, a concentração da produção de IA em grandes corporações privadas intensifica a falta de transparência e aprofunda desigualdades tecnológicas, especialmente em países periféricos. Conclui-se que a confiança na IA deve ser entendida como construção contínua e situada, entrelaçada a decisões humanas, interesses econômicos e estruturas de poder. Assim, confiar na IA exige manter a responsabilidade sobre seu uso, com mediação crítica e fundamentação em princípios de justiça, equidade e liberdade cognitiva, reforçando a necessidade urgente de uma educação crítica e da literacia em IA.

Palavras-chave: Inteligência artificial; Confiabilidade; Mediação humana; Literacia em IA.

Abstract

The Artificial Intelligence (AI), developed since the 1950s, is a general-purpose technology that requires critical analysis of its ethical, epistemological, and sociotechnical implications.

[1] Universidade Federal de Santa Catarina – UFSC. Araranguá/SC, Brasil. E-mail: juliana.pires@ufsc.br. Orcid: https://orcid.org/0009-0001-2663-535X
[2] Universidade Federal de Santa Catarina - UFSC. Florianópolis/SC, Brasil. E-mail: priscila.luz10661@edu.itajai.sc.gov.br. Orcid: https://orcid.org/0009-0005-5464-7461
[3] Universidade do Vale do Itajaí - UNIVALI. Itajaí/SC, Brasil. E-mail: rafaelarlindo@hotmail.com. ORCID: https://orcid.org/0000-0002-7644-1684

While it promises innovation and efficiency, it raises concerns such as algorithmic biases, social bubbles, hallucinations, and model opacity. In this theoretical essay, AI reliability is discussed from three main perspectives: biases, "illusions," and the decisions made based on its outputs. A qualitative methodological approach is adopted, centered on a literature review and engaging with authors such as Rivoltella (2024) and Fantin (2024). It is shown that algorithmic biases reflect prejudices embedded in training data, as seen with ChatGPT and facial recognition systems that fail to identify black faces, reinforcing hegemonic cultures and affecting the educational field. The "illusions" contribute to the spread of misinformation and strengthen the post-truth era. In this context, AI literacy becomes essential so that users can recognize limitations and flaws. Automated decisions often rely on opaque models that compromise explainability and require active human mediation. Furthermore, the concentration of AI development within large private corporations exacerbates the lack of transparency and deepens technological inequalities, especially in peripheral countries. It is concluded that trust in AI should be understood as an ongoing and situated construction, intertwined with human decisions, economic interests, and power structures. Thus, trusting AI requires maintaining responsibility over its use, through critical mediation and grounding in principles of justice, equity, and cognitive freedom, reinforcing the urgent need for critical education and AI literacy**.**

Keywords - Artificial intelligence; Trustworthiness; Human mediation; AI literacy.

Resumen

La Inteligencia Artificial (IA), desarrollada desde la década de 1950, es una tecnología de propósito general que requiere un análisis crítico de sus implicaciones éticas, epistemológicas y sociotécnicas. Aunque promete innovación y eficiencia, suscita preocupaciones como sesgos algorítmicos, burbujas sociales, alucinaciones y opacidad de los modelos. En este ensayo teórico, la confiabilidad de la IA se discute a partir de tres ejes: los sesgos, las "ilusiones" y las decisiones tomadas con base en sus respuestas. Para ello, se adopta un enfoque metodológico cualitativo, centrado en la revisión de literatura, dialogando con autores como Rivoltella (2024) y Fantin (2024). Se muestra que los sesgos algorítmicos reflejan prejuicios presentes en los datos de entrenamiento, como ocurre con ChatGPT y sistemas de reconocimiento facial que fallan en la identificación de rostros negros, reforzando culturas hegemónicas y afectando el ámbito educativo. Las "ilusiones" contribuyen a la difusión de desinformación y fortalecen la posverdad. En este contexto, la alfabetización en IA se vuelve esencial para que los usuarios reconozcan limitaciones y fallos. Las decisiones automatizadas se basan con frecuencia en modelos opacos, que comprometen la explicabilidad y exigen mediación humana activa. Además, la concentración de la producción de IA en grandes corporaciones privadas intensifica la falta de transparencia y profundiza las desigualdades tecnológicas, especialmente en países periféricos. Se concluye que la confianza en la IA debe entenderse como una construcción continua y situada, entrelazada con decisiones humanas, intereses económicos y estructuras de poder. Así, confiar en la IA exige mantener la responsabilidad sobre su uso, con mediación crítica y fundamentación en principios de justicia, equidad y libertad cognitiva, reforzando la necesidad urgente de una educación crítica y de la alfabetización en IA.

Palabras clave - Inteligencia artificial; Confiabilidad; Mediación humana; Alfabetización en IA.

Introdução

A Inteligência Artificial (IA) teve origem na década de 1950, consolidando-se como um campo da ciência da computação. Hoje, diversas aplicações demonstram o potencial da IA nas áreas de economia, saúde, negócios, cultura e nas relações sociais (Santaella, 2023). Sua pervasividade e as múltiplas vicissitudes que produz nos contextos sociais, econômicos e educacionais fazem com que ela seja considerada uma tecnologia de propósito geral. Essa condição demanda olhares críticos e interdisciplinares diante de suas implicações epistemológicas (relacionadas à produção da verdade e à opacidade dos sistemas), éticas (como a privacidade, os vieses e a discriminação) e sociotécnicas (que envolvem desigualdade, dependência tecnológica e exclusão digital).

Se, por um lado, a IA prenuncia a personalização, automatização, eficiência e inovação, por outro, suscita preocupações quanto aos vieses algorítmicos, à formação de bolhas sociais, às alucinações informacionais e à opacidade dos modelos. A confiabilidade da IA é basilar no debate educativo, revelando tensões entre técnica e conhecimento, tradições e inovações, riscos e possibilidades. Nesse campo de ambivalências, torna-se urgente refletir: até que ponto é possível confiar na IA?

Para problematizarmos essa resposta, optamos por adotar o formato de ensaio teórico, entendido como uma produção textual de caráter formal e argumentativo, que privilegia a reflexão crítica e interpretativa. Tal escolha permite ao autor maior liberdade para articular ideias, levantar hipóteses e construir sentidos a partir de diferentes perspectivas (Severino, 2007).

Neste ensaio, nos propomos a discutir, em três frentes complementares, os limites da confiabilidade cognitiva da IA: (1) os vieses algorítmicos que atravessam seus processos; (2) as “ilusões” produzidas por respostas incorretas ou enganosas; e (3) as decisões que se tomam com base nessas respostas. Para tanto, adota-se uma abordagem metodológica qualitativa, centrada na revisão de literatura.

A fundamentação teórica está ancorada primariamente em materiais acadêmicos e nas discussões promovidas durante as aulas da disciplina *Educação, Mídia e Inteligência Artificial: navegar na paisagem cultural contemporânea*, do Programa de Pós-Graduação em Educação da Universidade Federal de Santa Catarina, realizada no primeiro semestre de 2025,

complementada por uma investigação exploratória que buscou aprofundar a compreensão sobre IA.

Vieses algorítmicos: conceito e implicações

A noção de "viés" remete, originalmente, a uma inclinação, uma distorção em relação à neutralidade. Quando aplicada aos algoritmos, essa noção assume contornos técnicos e socioculturais, pois se refere à tendência sistemática de uma IA apresentar resultados parciais ou discriminatórios. (Rossetti, Angeluci; 2021). Isso evidencia que os algoritmos não são neutros, pois reproduzem ou até amplificam discriminações presentes nos dados com os quais foram treinados, gerando "[...] decisões ou previsões carregadas de preconceito" (Correia, 2023, n.p.).

No caso do ChatGPT, Rivoltella (2024, p. 32) explica como as respostas do modelo podem ter vieses dependendo da base de dados treinada:

> [...] tem sido repetidamente apontado que ele [ChatGPT] responde com base no modelo de um norte-americano de 40 anos, branco, da costa leste e progressista. Assim, o que parece ser informação na resposta às nossas perguntas pode conter preconceitos de raça, crença religiosa, afiliação cultural, ou ser construído com base em estereótipos.

Silva (2020), ao investigar a visão computacional e o racismo algorítmico, exemplifica vieses discriminatórios identificados por usuários, nos quais a população negra é sistematicamente prejudicada. Um dos casos relatados envolve a pesquisadora Joy Buolamwini, que, ainda estudante no Georgia Tech, percebeu, em feiras de tecnologia, que os robôs de reconhecimento facial identificavam seus colegas brancos, mas não reconheciam seu próprio rosto negro. A pesquisadora realizou um experimento com uma máscara branca sem feições e, assim, foi reconhecida pelas máquinas. O caso deu origem ao documentário Coded Bias (Netflix), que evidencia como os sistemas de reconhecimento facial, treinados sobretudo com rostos brancos, comprometem a equidade tecnológica e falham no reconhecimento de rostos negros.

Na esfera da segurança pública, o uso de reconhecimento facial evidencia um viés de representação quando a tecnologia identifica com mais precisão rostos brancos do que negros ou indígenas. Isso ocorre devido à sub-representação desses grupos nos dados de treinamento.

ISSN: 2526-849X

Como alerta Cathy O'Neil (2020), os "algoritmos de destruição em massa" podem gerar sistemas que perpetuam desigualdades sob uma aparência de precisão matemática. Nesse contexto, os erros de identificação podem ter consequências graves, sobretudo para populações racializadas.

Como exemplo de situação real, cita-se uma publicação da deputada estadual Renata Souza (RJ), no Instagram, em 26 de dezembro de 2023. Na ocasião, a parlamentar utilizou uma IA para gerar uma imagem no estilo Pixar, inspirada na tendência de criar pôsteres da Disney (Araújo; Araújo, 2024, p. 100).

> A descrição fornecida por ela solicitava a representação imagética de "uma mulher negra, de cabelos afro, com roupas de estampa africana num cenário de favela" [...]. No entanto, a imagem gerada pela IA apresentou uma mulher com uma arma na mão, associando indevidamente sua identidade e o cenário de favela à violência. [...] a deputada expressou sua preocupação com o viés racista presente nas tecnologias, como o reconhecimento facial, e a necessidade de revisar essas tecnologias e procedimentos para garantir a segurança e a justiça para as pessoas negras.

A Figura 1 apresenta a imagem gerada, evidenciando como os algoritmos podem reproduzir e perpetuar desigualdades raciais, sociais e outras, conforme os dados em que foram treinados. Essa observação amplia o debate sobre o viés algorítmico e seus impactos nas estruturas de poder.

Figura 1. Imagem racista gerada por IA

Fonte: Araújo e Araújo (2024, p. 101).

ISSN: 2526-849X

Além dos vieses supracitados, a IA também pode gerar distorções ou reforçar culturas hegemônicas, invisibilizando a pluralidade de identidades culturais e linguísticas. Isso ocorre, em grande parte, devido ao uso de bancos de dados de treinamento compostos majoritariamente por textos em línguas dominantes, o que restringe a diversidade representada nos modelos. Como alertam Almeida et al. (2025, p. 14), esse processo pode reforçar "[…] narrativas neocoloniais, discriminações étnico-raciais, sexistas, dentre outros, limitando o alcance a uma gama diversificada de vozes e saberes".

No campo educacional, esses mesmos mecanismos de viés se manifestam em tecnologias como sistemas de recomendação, correção automatizada e predição de desempenho. Embora tais recursos possam parecer promissores, eles carregam o risco de reforçar preconceitos, excluir sujeitos e comprometer trajetórias formativas, sobretudo quando operam sobre dados marcados por desigualdades (Almeida et al., 2025).

No caso dos sistemas de avaliação automática, baseados em aprendizado de máquina, estes podem prejudicar estudantes que se expressam fora do padrão linguístico dominante, reproduzindo discriminações culturais e socioeconômicas. Rivoltella (2024) alerta que a IA, sem supervisão crítica, pode naturalizar decisões técnicas como se fossem neutras, ignorando o caráter ideológico de seus critérios. A percepção de uma "objetividade matemática" desses sistemas pode mascarar os vieses inerentes aos dados de treinamento e aos objetivos dos desenvolvedores. A IA apenas reproduzirá esse padrão, sob a aparência de eficiência. Nesse sentido, os algoritmos não apenas medem ou classificam: eles intervêm e moldam a experiência educativa, atuando como mediadores sociotécnicos com implicações éticas, pedagógicas e políticas. Por isso, a necessidade de mediação humana para desvelarmos suas opacidades atuando de forma reflexiva, ética e crítica.

Sayad (2023, p. 82) fala da complexidade ligada à ética da IA.

> Portanto, a complexidade ligada à ética da IA, sobretudo no que diz respeito à técnica de redes neurais profundas, envolve muitos aspectos: subjetividade humana (de desenvolvedores, implementadores e usuários intermediários), quantidade e qualidade das bases de dados, mas também efeitos da opacidade intrínseca à própria técnica.

Por mais complexas e opacas que sejam as técnicas envolvidas, é fundamental atribuir aos seres humanos a responsabilidade pelos problemas éticos e vieses gerados pelos sistemas

ISSN: 2526-849X

de IA. Se valores não forem incorporados ao processo de treinamento dos dados e à definição dos objetivos desses sistemas, arriscamos intensificar preconceitos e aprofundar desigualdades sociais já existentes (Harari, 2024).

Diante disso, torna-se urgente incluir, na formação de profissionais da área de IA, temas relacionados à justiça social, à ética e à equidade, capacitando-os para reconhecer e enfrentar os vieses. É igualmente necessário garantir a representação e a participação efetiva de grupos historicamente subalternizados, como mulheres, negros e povos indígenas, nos espaços de desenvolvimento tecnológico. A formação de profissionais conscientes e diversos é um passo essencial para que a IA possa, de fato, contribuir com a inovação e a promoção da equidade no campo educacional (Almeida et al., 2025).

Ilusão: alucinações, pós-verdade e bolhas sociais

A tríade conceitual que sustenta esta seção articula fenômenos de naturezas distintas, mas com efeitos convergentes sobre a percepção da realidade. Enquanto as alucinações derivam da natureza probabilística da IA generativa — produzindo erros factuais sob uma roupagem de plausibilidade linguística —, a desinformação opera na esfera da intencionalidade, mobilizando ecossistemas de circulação para distorcer o debate público. Esse cenário é agravado pelas bolhas sociais, cujos sistemas de recomendação promovem um fechamento informacional que valida a pós-verdade e isola o sujeito de perspectivas divergentes. No contexto digital e hiperconectado, a IA, por meio de seus algoritmos, filtra e direciona escolhas, vigia, cria bolhas informativas e contribui para o capitalismo digital e a disseminação de *fake news*. É fundamental compreender conceitos como metaverdade, que alude à construção de "verdades" baseadas em narrativas subjetivas e dissimuladas, e pós-verdade, em que crenças pessoais ganham mais peso que fatos objetivos, sendo esta compreendida como o "conjunto de fatos ou informações que, sem fundamento e propagados de maneira repetitiva, são tidos como verdadeiros" (Dicio, [s.d.], p. 1). Assim, fomentar o pensamento crítico, capacitando os estudantes para autenticar e curar informações, revela-se um ato fundamental na construção do conhecimento. Divergir e verificar são ações essenciais, sob o risco de, como alertam Blum e Endo (2023, p. 79), nos tornarmos "incapazes de constituir uma única experiência senão a de dispensar as próprias experiências".

Sob essa ótica, Velander et al. (2023), ao investigarem o uso da IA na educação, expressam preocupações quanto ao tratamento de dados em mídias sociais e à sua influência nos comportamentos individuais. Destacam, em especial, o fenômeno das bolhas sociais, que limita o pensamento crítico ao expor os sujeitos apenas a conteúdos que confirmam suas visões prévias, dificultando a capacidade de questionar, habilidade central para a emancipação. Ainda no contexto educacional, uma implicação está no uso, por alunos, de sistemas generativos para apoiar um trabalho, o que pode induzi-los a acreditar na confiabilidade do texto gerado, sem perceberem que estão diante de uma estrutura semântica apenas plausível, mas não necessariamente correta do ponto de vista factual. A IA, nesse processo, pode errar ao responder, produzindo as chamadas "alucinações" (Sampaio et al., 2024). Rivoltella (2024) adverte que a literacia da IA deve incluir não apenas a capacidade de formular *prompts,* mas também de reconhecer limites e falhas do sistema, interrogando criticamente a veracidade de suas saídas/respostas.

Fantin (2024) destaca a importância de articular o digital e o analógico, o natural e o artificial, ao evidenciar como o ambiente circundante contribui para o desenvolvimento da criatividade e para o avanço cognitivo. A busca por espaços híbridos configura-se como uma alternativa significativa para os processos de ensino e aprendizagem em cenários conectados. Tal proposta precisa da mediação docente, uma vez que baseia-se em uma "[...] educação dialógica que deve estar situada na cultura, na linguagem, na política e na vida de estudantes e professores de modo a transcender a experiência subjetiva, particular e local para expandir-se em uma dimensão mais ampla" (Fantin, 2024, p. 96).

Nessa perspectiva, reconhecer os desafios contemporâneos torna-se fundamental, especialmente diante dos usos inadequados das tecnologias e da disseminação de desinformação, em que erros são intencionalmente inseridos nos meios de comunicação com a finalidade de influenciar o consumo, a política, entre outros âmbitos sociais. Reforça-se, aqui, o papel do professor como curador de conteúdo, capaz de mediar criticamente a seleção de materiais confiáveis e de qualificar o acesso à informação. Desenvolver essa competência é um dos maiores desafios da cultura digital contemporânea, marcada, como diria Norman (1998), pela invisibilidade das mediações. É imperativa a promoção de uma cultura de cibersegurança e a sensibilização para o uso ético e responsável das tecnologias educacionais e IA.

ISSN: 2526-849X

Decisão: inferências algorítmicas e mediação humana

A decisão, como ato humano, envolve ponderação, intencionalidade e responsabilização. Contudo, quando mediada por algoritmos, ela passa a ser influenciada por inferências estatísticas produzidas a partir de correlações, e não necessariamente por causalidades compreendidas. Como mostram Panciroli e Rivoltella (2024), as decisões algorítmicas são tomadas com base em modelos opacos, cujos critérios de funcionamento muitas vezes não são acessíveis nem aos próprios programadores. Esse fenômeno é conhecido como caixa-preta algorítmica, e compromete o princípio de explicabilidade, condição fundamental para a legitimidade de qualquer decisão em contextos éticos, legais e educacionais.

Na educação, as decisões automatizadas também têm ganhado espaço, especialmente no uso de plataformas adaptativas e sistemas de análise de desempenho. Embora úteis para personalização da aprendizagem, esses sistemas podem induzir decisões sobre trajetórias educacionais com base em métricas que não captam aspectos qualitativos da experiência do estudante, como criatividade, pensamento divergente ou contexto sociocultural. Rivoltella (2024) chama a atenção para a importância de não confundir performance com aprendizagem, nem reduzir a educação a métricas tecnicamente eficientes.

A promoção de uma comunicação transparente e a tomada de decisões informadas requerem uma formação holística aos educadores. Nesse contexto, é fundamental oferecer formação contínua sobre as capacidades e limitações da IA, de modo a subsidiar decisões éticas baseadas em dados. Tais formações devem considerar que o processo de ensinar exige atualização constante, pois, como afirma Freire (1996, p. 12) “quem ensina aprende ao ensinar e quem aprende ensina ao aprender”. Os processos educativos não se reduzem à transmissão de saberes, mas envolvem práticas dialógicas que consideram contextos, mediações e transformações sociais. Nesse sentido, a literacia em IA torna-se essencial como parte de uma educação voltada para a cultura digital crítica.

> Falar da dimensão da literacia significa referir-se às linguagens: não se trata apenas de habilidades em informática ou escrita de código; trata-se também do léxico da IA que precisa ser desenvolvido, da compreensão de como ela funciona, tanto na frente quanto nos bastidores das interfaces (Ng et al., 2021 apud Panciroli; Rivoltella, 2024, p. 33).

ISSN: 2526-849X

A presença da IA exige atenção, incluindo o domínio de seus fundamentos, o uso crítico de suas ferramentas, a avaliação e a criação a partir de seus recursos, bem como a reflexão sobre aspectos éticos fundamentais, como justiça, responsabilidade, transparência e segurança. Tal necessidade torna-se fundamental diante da concentração de sua produção e difusão em grandes corporações privadas (*Big Techs)*, que detêm vasto poder de investimento e influência global. Tais empresas justificam frequentemente a ausência de transparência nas decisões automatizadas sob a alegação de proteção do segredo comercial, o que gera zonas de opacidade e levanta importantes questões éticas relacionadas aos algoritmos e aos conjuntos de dados utilizados nos processos de treinamento (Sampaio et al., 2024).

Na contramão do monopólio e das possibilidades de recursos de IA do norte global, no setor público e em países em desenvolvimento, como o Brasil, o uso de ferramentas gratuitas, com funcionalidades limitadas, evidencia desigualdades de acesso e aprofunda as disparidades tecnológicas. Tais restrições impactam diretamente a aplicabilidade dessas tecnologias no cotidiano escolar, suscitando questões sobre regionalidade, vieses algorítmicos, ética e representatividade cultural, o que pode culminar no "acirramento das exclusões em relação às representatividades" (Almeida et al., 2025). Aqui emergem questões cruciais: como preparar educadores para o uso crítico e responsável dessas tecnologias, de modo a mitigar riscos relacionados à segurança, privacidade, desinformação e discriminação? E ainda, como evitar que a IA se torne um vetor de ampliação das desigualdades socioeconômicas e culturais no campo educacional?

Considerações finais

A reflexão em torno da confiabilidade cognitiva da IA, com base nos vieses, nas ilusões e nas decisões mediadas por algoritmos, revela um cenário complexo e ambíguo. De um lado, a IA apresenta-se como um poderoso instrumento de apoio à ação humana, capaz de otimizar processos, gerar conhecimento e ampliar horizontes de interação. De outro, sua lógica de funcionamento, ancorada em dados parciais, opacidades estruturais e ausência de compreensão semântica, impõe limites à sua confiabilidade enquanto fonte de conhecimento e tomada de decisão.

A confiabilidade da IA não pode ser assumida como um dado, mas como uma construção contínua, situada e negociada. A IA, enquanto artefato técnico e cultural, é permeada por decisões humanas, interesses econômicos e estruturas de poder que determinam seu funcionamento e impacto. Podemos confiar na IA até o ponto em que compreendemos seus limites, desvelamos suas opacidades e mantemos ativa a mediação humana. A confiança não deve ser cega, mas reflexiva; nem ser delegada, mas compartilhada. Trata-se de educar para a criticidade e ética no uso de IA. Nesse cenário, autores como Rivoltella e Panciroli (2024) propõem um caminho centrado na educação crítica, na literacia em IA e na mediação consciente dos processos tecnológicos.

A resposta à pergunta que intitula este ensaio, portanto, é condicional: podemos confiar na IA na medida em que não abrimos mão da nossa responsabilidade diante dela. A IA pode ser uma aliada na construção do conhecimento, desde que seja moldada por princípios de justiça, equidade, sustentabilidade e liberdade cognitiva.

Referências

ALMEIDA, Ana Paula; ARAÚJO, Cláudia Helena dos Santos; FERRARO, Danielle Soares e Silva Bicudo; VIEIRA, Lívia Carolina; CASTRO, Karolina Batista; COELHO, Márcia Azevedo; QUADROS, Paulo da Silva. **Carta de recomendação para o uso da inteligência artificial na educação: desafios e potencialidades.** 1. ed. [S. l.]: Instituto Federal de Educação, Ciência e Tecnologia de Goiás, 2025. e-book, 1,55 MB. Disponível em: http://educapes.capes.gov.br/handle/capes/972722. Acesso em: 19 jun. 2025.

ARAÚJO, Júlio; ARAÚJO, Júlio. Racismo algorítmico e inteligência artificial: uma análise crítica multimodal. **Revista Linguagem em Foco**, v.16, n.2, 2024. p. 89-109. Disponível em: https://revistas.uece.br/index.php/linguagememfoco/article/view/13108 . Acesso em: 20 de jun. 2025.

BLUM, Rodrigo; ENDO, Paulo Cesar. O quarto golpe. **Percurso**, São Paulo, Brasil, v. 35, n. 70, p. 69–80, 2023. Disponível em: https://percurso.openjournalsolutions.com.br/index.php/ojs/article/view/1427. Acesso em: 2 jul. 2025.

CORREIA, Ana-Paula. É o ChatGPT uma nova tendência no Ensino Superior? Notícias, **Revista Docência e Cibercultura.** Abril de 2023. Online. Disponível em: https://www.e-publicacoes.uerj.br/index.php/re-doc/an nouncement/view/1622.
Acesso em: 19 jun. 2025.

DICIO. Pós-verdade. Dicionário *Online* de Português, [s.d.]. Disponível em: https://www.dicio.com.br/pos-verdade/. Acesso em: 1 jul. 2025.

FANTIN, Mônica. Confinamentos, pedagogia do lugar e terceiro espaço na educação. In: BORGES, Gabriela; GRÁCIO, Rui Alexandre; RIBEIRO, Orquídea (Orgs.). **Cidadania digital e culturas do contemporâneo.** Coimbra: Grácio Editor, 2024. p. 93–102. ISBN 978-989-35413-3-3. DOI: 10.5281/zenodo.11127615. Disponível em: https://ciac.pt/wp-content/uploads/2024/07/Cidadania-digital-e-culturas-do-contemporaneo_digital-1.pdf. Acesso em: 20 jun. 2025.

FREIRE, Paulo. **Pedagogia da Autonomia: saberes necessários à prática educativa.** São Paulo: Paz e Terra, 1996.

HARARI, Yuval Noah. **Nexus:** uma breve história das redes de informação, da Idade da Pedra à inteligência artificial. 1. ed. São Paulo: Companhia das Letras, 2024. 504 p.

NORMAN, Donald. A. **The Invisible Computer:** Why Good Products Can Fail, the Personal Computer Is So Complex, and Information Appliances Are the Solution. Cambridge, MA: MIT Press, 1998.

O´NEIL, Cathy. **Algoritmos de destruição em massa:** como o big data aumenta a desigualdade e ameaça a democracia. Santo André, SP: Rua do Sabão, 2020.

PANCIROLI, Paolo; RIVOLTELLA, Pier Cesare. Collaborating with Machines. AI, Literacies, School. In: **Artificial Intelligence in Education.** Tradução de SCHOLÉ – Grupo de Estudos e Pesquisas sobre Inovação na Educação. 2024. Disponível em: https://www.morcelliana.net/img/cms/Rivista%20Schol%C3%A9/2024/1%202024/Panciroli%20Rivoltella%20Schole%CC%81%201_2024%20finale-15-48.pdf. Acesso em: 22 jun. 2025.

RIVOLTELLA, Pier Cesare. Talking to Machines: Semiotic Analysis, Implications for Teaching and Media Literacy. *AN-ICON*. **Studies in Environmental Images**. 2024 [ISSN 2785-7433], 3(II), 17–35. https://doi.org/10.54103/ai/23944

ROSSETTI, Regina; ANGELUCI, Alan. Ética Algorítmica: questões e desafios éticos do avanço tecnológico da sociedade da informação. **Galáxia,** São Paulo, 2021. https://doi.org/10.1590/1982-2553202150301. Disponível em: https://www.scielo.br/j/gal/a/R9F45HyqFZMpQp9BGTfZnyr . Acesso em: 19 jun. de 2025.

SANTAELLA, Lucia. **A inteligência artificial é inteligente?** São Paulo: Edições 70, 2023.

SAMPAIO, Rafael Cardoso; NICOLÁS, Maria Alejandra; JUNQUILHO, Tainá Aguiar; SILVA, Luiz Rogério Lopes; FREITAS, Christiana Soares de; TELLES, Márcio; TEIXEIRA, João Senna; ESCÓSSIA, Fernanda da; SANTOS, Luiza Carolina dos. ChatGPT e outras IAs transformarão a pesquisa científica: reflexões sobre seus usos. **Revista de Sociologia e Política**, [S.L.], v. 32, p. 1-24, 2024. FapUNIFESP (SciELO). http://dx.doi.org/10.1590/1678-98732432e008. Acesso em: 04 ago. 2025.

SAYAD, Alexandre Le Voice. **Inteligência Artificial e Pensamento Crítico:** caminhos para uma educação midiática. 1.ed. São Paulo: Instituto Palavra Aberta, 2023.

SEVERINO, Antônio Joaquim. **Metodologia do Trabalho Científico**. 23. ed. São Paulo: Cortez, 2007.

SILVA, Tarcízio. Visão Computacional e Racismo Algorítmico: Branquitude e Opacidade no Aprendizado de Máquina. **Revista ABPN**, v. 12, p. 428-448, 2020. Disponível em: https://abpnrevista.org.br/site/article/view/744. Acesso em: 19 jun. 2025.

VELANDER, Johanna; TAIYE, Mohammed Ahmed; OTERO, Nuno; MILRAD, Marcelo. Artificial Intelligence in K-12 Education: eliciting and reflecting on swedish teachers' understanding of ai and its implications for teaching & learning. **Education And Information Technologies**, [S.L.], v. 29, n. 4, p. 4085-4105, 3 jul. 2023. Springer Science and Business Media LLC. http://dx.doi.org/10.1007/s10639-023-11990-4.

Recebido: setembro/2025
Aprovado: fevereiro/2026