

Inteligência artificial e avaliação escolar: uma análise crítica dos vieses algorítmicos

Márcio Silveira Nascimento¹ 

Sidney Pires Martins² 

Aline Santos Mossette³ 

Éber José dos Santos⁴ 

Resumo

A crescente incorporação da Inteligência Artificial (IA) nos sistemas educacionais tem transformado práticas avaliativas, oferecendo potencial para personalização, eficiência e apoio à gestão pedagógica. No entanto, a presença de vieses algorítmicos pode gerar avaliações injustas, especialmente para grupos historicamente marginalizados. O presente estudo constituiu-se como uma investigação teórico-reflexiva, fundamentada em revisão bibliográfica interpretativa, com o objetivo de examinar criticamente os vieses algorítmicos na avaliação escolar, discutir suas implicações éticas e pedagógicas e propor estratégias de mitigação. A análise indicou que distorções podem surgir de dados históricos desbalanceados, da arquitetura inadequada de algoritmos e das limitações na adaptação às necessidades individuais dos estudantes. Os resultados evidenciam que o uso ético e eficaz da IA depende da integração entre tecnologia e mediação humana, destacando o papel insubstituível do professor como mediador crítico e contextualizador dos resultados. Conclui-se que, embora a IA ofereça oportunidades significativas para personalização do ensino, sua implementação requer atenção à justiça social, à diversidade cultural e à equidade educacional. Assim, este estudo contribui para orientar pesquisadores, gestores e educadores na construção de práticas pedagógicas mais inclusivas, justas e contextualizadas, promovendo uma educação reflexiva e equitativa no uso de sistemas inteligentes.

Palavras-chave: inteligência artificial; vieses algorítmicos; tecnologia educacional.

¹ Doutorando e Mestre em Ensino Tecnológico, pelo Instituto Federal do Amazonas - IFAM. Pesquisador do Grupo de Estudos e Pesquisas em Políticas Públicas e Formação de Profissionais da Educação - GEPPFOR/UFV. Manaus, AM, Brasil. ORCID: <https://orcid.org/0000-0002-8963-3140>. E-mail: marciosn.geo@gmail.com

² Doutorando em Políticas Sociais, pela Universidade Federal de Viçosa - UFV. Pesquisador do Grupo de Estudos e Pesquisas em Políticas Públicas e Formação de Profissionais da Educação - GEPPFOR/UFV. Viçosa, MG, Brasil. ORCID: <https://orcid.org/0000-0002-4890-9307>. E-mail: sidney.martins@ufv.br

³ Doutoranda em Difusão do Conhecimento, pelo Instituto Federal da Bahia - IFBA. Pesquisadora do Grupo de Estudos e Pesquisas em Políticas Públicas e Formação de Profissionais da Educação - GEPPFOR/UFV. Salvador, BA, Brasil. ORCID: <https://orcid.org/0000-0003-0271-2522>. E-mail: alinemossette1@gmail.com

⁴ Doutor e Mestre em Língua Portuguesa, pela Pontifícia Universidade Católica de São Paulo - PUC/SP. Pesquisador do Grupo de Estudos e Pesquisas em Políticas Públicas e Formação de Profissionais da Educação - GEPPFOR/UFV. São Paulo, SP, Brasil. ORCID: <https://orcid.org/0000-0003-2021-2062>. E-mail: ejsantos2010@gmail.com

Artificial intelligence and school assessment: a critical analysis of algorithmic biases

Abstract

The growing incorporation of Artificial Intelligence (AI) into educational systems has transformed assessment practices, offering potential for personalization, efficiency, and support for pedagogical management. However, the presence of algorithmic biases can lead to unfair assessments, especially for historically marginalized groups. This study was a theoretical-reflective investigation, based on an interpretive literature review, with the aim of critically examining algorithmic biases in school assessment, discussing their ethical and pedagogical implications, and proposing mitigation strategies. The analysis indicated that distortions can arise from unbalanced historical data, inadequate algorithm architecture, and limitations in adapting to students' individual needs. The results show that the ethical and effective use of AI depends on the integration of technology and human mediation, highlighting the irreplaceable role of the teacher as a critical mediator and contextualizer of results. It is concluded that, although AI offers significant opportunities for personalizing teaching, its implementation requires attention to social justice, cultural diversity, and educational equity. Thus, this study contributes to guiding researchers, managers, and educators in the construction of more inclusive, fair, and contextualized pedagogical practices, promoting reflective and equitable education in the use of intelligent systems.

Keywords: artificial intelligence; algorithmic biases; educational technology.

Inteligencia artificial y evaluación escolar: un análisis crítico de los sesgos algorítmicos

Resumen

La creciente incorporación de la Inteligencia Artificial (IA) a los sistemas educativos ha transformado las prácticas de evaluación, ofreciendo posibilidades de personalización, eficiencia y apoyo a la gestión pedagógica. Sin embargo, la presencia de sesgos algorítmicos puede conducir a evaluaciones injustas, especialmente para grupos históricamente marginados. Este estudio fue una investigación teórica y reflexiva, basada en una revisión interpretativa de la literatura, con el objetivo de examinar críticamente los sesgos algorítmicos en la evaluación escolar, debatir sus implicaciones éticas y pedagógicas y proponer estrategias de mitigación. El análisis indicó que las distorsiones pueden surgir de datos históricos desequilibrados, una arquitectura inadecuada de los algoritmos y limitaciones en la adaptación a las necesidades individuales de los alumnos. Los resultados muestran que el uso ético y eficaz de la IA depende de la integración de la tecnología y la mediación humana, destacando el papel insustituible del profesor como mediador crítico y contextualizador de los resultados. Se concluye que, aunque la IA ofrece importantes oportunidades para personalizar la enseñanza, su aplicación requiere prestar atención a la justicia social, la diversidad cultural y la equidad educativa. Así, este estudio contribuye a orientar a investigadores, gestores y educadores en la construcción de prácticas pedagógicas más inclusivas, justas y contextualizadas, promoviendo una educación reflexiva y equitativa en el uso de sistemas inteligentes.

Palabras clave: inteligencia artificial; sesgos algorítmicos; tecnología educativa.

Introdução

Os algoritmos influenciam nossas vidas de diversas maneiras e, muitas vezes, são percebidos como instrumentos inerentemente justos e objetivos (Lee, 2018). No campo educacional, a crescente presença da IA transforma as práticas de ensino e aprendizagem, ampliando possibilidades que até pouco tempo pareciam distantes.

Atualmente, esses sistemas são empregados em diferentes contextos

escolares e acadêmicos (Silva; Costa, 2024). Eles podem analisar grandes volumes de dados educacionais para oferecer recomendações personalizadas, adaptando conteúdos, atividades e ritmos às necessidades individuais dos estudantes (Luckin *et al.*, 2016). Esse processo, conhecido como aprendizagem personalizada, representa um avanço importante, pois reconhece que os alunos não aprendem de forma homogênea e que os processos de ensino podem ser mais eficazes quando consideram as singularidades cognitivas, culturais e socioemocionais.

Além disso, ferramentas baseadas em IA têm sido aplicadas para fornecer *feedbacks* imediatos, o que permite aos estudantes compreender seus erros em tempo real e ajustem suas estratégias de aprendizagem (Holmes; Bialik; Fadel, 2019). Outros exemplos incluem sistemas de tutoria inteligente, plataformas de recomendação de conteúdos, análise preditiva de desempenho acadêmico e avaliação automatizada de tarefas, que auxiliam professores na gestão da sala de aula e na personalização das experiências de aprendizagem. Dessa forma, embora a IA ofereça oportunidades promissoras para tornar a educação mais adaptativa e inclusiva, é fundamental considerar também os vieses que podem surgir desses algoritmos e suas implicações éticas e pedagógicas.

Outro aspecto relevante no uso de tecnologias inteligentes na educação é o suporte oferecido ao trabalho docente. Sistemas de correção automática de provas, plataformas de gestão de desempenho escolar e assistentes virtuais reduzem a sobrecarga de tarefas administrativas e permitem que os professores direcionem mais tempo ao acompanhamento pedagógico e ao desenvolvimento integral dos estudantes (Zawacki-Richter *et al.*, 2019). Além disso, esses recursos possibilitam a identificação precoce de dificuldades de aprendizagem por meio da análise de padrões comportamentais e cognitivos, o que favorece a implementação de estratégias de intervenção mais precisas e oportunas.

Por outro lado, é importante destacar que o uso da IA na educação não se limita apenas à otimização de processos técnicos. Ele também redefine os modos de interação entre estudantes, professores e o próprio conhecimento (Menta; Brito, 2024). A introdução de tutores inteligentes, *chatbots* educacionais e plataformas adaptativas cria novas formas de mediação pedagógica, nas quais a tecnologia assume o papel de facilitadora do acesso ao saber. No entanto, esses avanços suscitam reflexões sobre o risco de uma dependência tecnológica excessiva, que



pode reduzir a autonomia crítica dos sujeitos e reforçar a lógica de mercantilização do ensino (Selwyn, 2022).

Contudo, como destaca Casagrande (2019), a introdução dessas tecnologias no contexto pedagógico exige um repensar profundo do papel do professor. Embora a IA possa automatizar tarefas rotineiras, como a correção de provas e o gerenciamento de dados, ela não substitui a dimensão humana do ensino, indispensável para promover o desenvolvimento crítico, ético e socioemocional dos estudantes.

Dessa forma, a IA deve ser compreendida como um recurso complementar que amplia as possibilidades do trabalho docente, mas não o substitui. O professor permanece como mediador fundamental, capaz de interpretar os dados gerados pelos sistemas inteligentes e contextualizá-los em práticas educativas que visem não apenas à eficiência, mas também à formação integral dos sujeitos.

No mesmo horizonte analítico, Heggler, Szmoski e Miquelin (2025) questionam se essa tecnologia, embora promova novas formas de interação, personalização do ensino e aceleração dos processos avaliativos, é capaz de assegurar tais benefícios sem comprometer a equidade e a eficácia educativa diante das distorções produzidas por modelos algorítmicos. Essa preocupação dialoga diretamente com o alerta de Noble (2018), que evidencia como sistemas automatizados tendem a reproduzir preconceitos estruturais presentes nos dados. No contexto da avaliação escolar, tal risco torna-se particularmente crítico, pois interfere de modo significativo nas trajetórias acadêmicas e sociais dos estudantes.

Adicionalmente, Tang e Su (2024) ressaltam que, embora modelos computacionais avancem na oferta de percursos personalizados de aprendizagem, eles permanecem limitados pela incapacidade de captar plenamente a complexidade e a variabilidade das necessidades discentes. Essa limitação estrutural evidencia que tais sistemas não substituem a análise pedagógica humana e requerem uso criterioso. Assim, a tecnologia ocupa uma posição ambígua no campo educacional, ao mesmo tempo em que apresenta potencial para reconfigurar práticas de ensino, também pode reproduzir e intensificar desigualdades já existentes caso seja aplicada sem mediação crítica e atenção às condições socioculturais que moldam a aprendizagem.

Diante desse cenário, o presente estudo caracterizou-se como uma investigação teórico-reflexiva, fundamentada em revisão bibliográfica interpretativa,



compreendida como análise crítica de obras teóricas relevantes. Conforme Lima e Miotto (2007), a revisão bibliográfica deve ir além da mera descrição de conteúdos, assumindo caráter problematizador e crítico. Esse tipo de abordagem possibilita a elaboração de um texto sistematizado, no qual as ideias extraídas da literatura são articuladas e analisadas à luz das interpretações do pesquisador. De forma complementar, Sousa, Oliveira e Alves (2021) ressaltam que a revisão bibliográfica interpretativa visa identificar, analisar e criticar documentos previamente publicados, contribuindo para a atualização e o aprimoramento do conhecimento existente.

A adoção desse método mostrou-se adequada aos objetivos da pesquisa, pois permitiu analisar de forma crítica as origens, os impactos e as possibilidades de enfrentamento das distorções produzidas por sistemas automatizados na avaliação escolar, contribuindo para o aprofundamento das dimensões éticas e pedagógicas envolvidas. Assim, o artigo propõe uma reflexão crítica sobre os vieses algorítmicos na avaliação escolar, examinando impactos, discutindo implicações éticas e pedagógicas e apresentando estratégias de mitigação, com o propósito de promover um uso mais justo, inclusivo e humano da IA na educação.

Compreendendo os Vieses Algorítmicos e suas Implicações na Educação

As técnicas de aprendizado de máquina representam uma ruptura significativa em relação às abordagens tradicionais da IA, pois, ao contrário dos métodos que dependem da programação explícita de regras complexas, permitem que a própria máquina desenvolva soluções de forma autônoma e adaptativa (Villani, *et al.*, 2018).

Essa capacidade de autodescoberta e ajuste contínuo amplia consideravelmente as possibilidades de aplicação da IA, desde sistemas de recomendação de conteúdos até plataformas de avaliação adaptativa, tornando o ensino mais personalizado e responsivo às necessidades individuais dos alunos. No entanto, essa autonomia algorítmica também introduz desafios inéditos, especialmente no que se refere à imparcialidade, à transparência e à confiabilidade das decisões automatizadas.

Entre os principais desafios, destacam-se os vieses algorítmicos, entendidos como distorções ou preconceitos incorporados em sistemas de IA que podem afetar diretamente decisões, análises e recomendações (Noble, 2018; Tang; Su, 2024). Eles têm origem em múltiplos fatores, sendo o mais evidente o uso de dados históricos



para treinar os algoritmos. Quando essas bases de dados carregam desigualdades sociais, estereótipos ou erros de classificação, o sistema tende a reproduzir e, em muitos casos, amplificar essas distorções de forma não intencional. Suresh e Guttag (2021) explicam que esses vieses funcionam como potenciais geradores de danos, pois introduzem características não planejadas ou prejudiciais nos dados, capazes de produzir resultados considerados indesejáveis ou socialmente problemáticos.

Além das distorções presentes nos dados, outros desvios podem emergir do próprio *design* algorítmico, como a atribuição inadequada de pesos a determinados fatores ou o uso de variáveis substitutas (*proxies*) que não representam fielmente as informações originais (Vieira, 2025). Esses elementos demonstram que as distorções produzidas por sistemas algorítmicos não se limitam a desafios de natureza técnica, mas configuram fenômenos profundamente atravessados por dimensões sociais, culturais e éticas. Como argumentam Noble (2018) e Tang e Su (2024), tecnologias aparentemente neutras podem reproduzir desigualdades históricas e reforçar mecanismos de exclusão, o que torna indispensável uma análise crítica e situada de seus modos de implementação.

No campo educacional, os efeitos dessas distorções podem assumir caráter amplo e complexo. Ferramentas de avaliação automatizada, sistemas de recomendação de conteúdos e mecanismos de monitoramento do desempenho escolar têm potencial para privilegiar determinados perfis de estudantes enquanto desfavorecem outros, sobretudo aqueles pertencentes a grupos historicamente marginalizados (Vieira, 2025). Ademais, o uso de informações de desempenho passado para prever desempenhos futuros tende a cristalizar assimetrias socioeconômicas, restringindo possibilidades formativas e perpetuando barreiras estruturais. Nesse cenário, a IA deixa de atuar apenas como ferramenta de otimização pedagógica e passa a exercer papel determinante na mediação das trajetórias educacionais, com implicações éticas, políticas e pedagógicas de grande relevância.

Diante desse cenário, torna-se evidente que o uso da IA na educação requer não apenas a compreensão técnica de algoritmos e dados, mas também a reflexão crítica sobre os impactos sociais dessas tecnologias. A supervisão contínua por parte de professores e gestores, a revisão constante das bases de dados e a implementação de estratégias de mitigação são essenciais para garantir que os sistemas de IA contribuam para a equidade, a justiça e a promoção de uma



aprendizagem inclusiva (Noble, 2018; Vieira, 2025). Assim, a reflexão sobre tendências algorítmicas que podem gerar distorções torna-se um componente central para o desenvolvimento de práticas pedagógicas éticas e para a construção de ambientes educativos verdadeiramente democráticos, nos quais a tecnologia atua como aliada e não como agente de perpetuação de desigualdades.

Impactos dos Vieses na Avaliação Escolar: Implicações Éticas e Pedagógicas

A utilização de IA na educação tem se consolidado como uma tendência global. Entretanto, sua incorporação não está isenta de dilemas éticos, pedagógicos e legais. Selwyn (2019) observa que a introdução de tecnologias no campo educacional frequentemente reforça desigualdades preexistentes, funcionando mais como um recurso que amplia oportunidades para os que já possuem acesso privilegiado do que como um mecanismo capaz de superar barreiras estruturais que excluem grupos marginalizados. Essa análise evidencia que a adoção acrítica de tecnologias digitais não assegura equidade; ao contrário, pode aprofundar assimetrias históricas quando não acompanhada de políticas e práticas efetivas de mitigação de desigualdades.

Nesse mesmo horizonte, destaca-se que, quando mal projetada ou implementada de forma inadequada, tende a reproduzir desigualdades em vez de combatê-las. Isso ocorre porque os algoritmos aprendem a partir de dados que frequentemente contêm vieses, replicando distorções no processo de aprendizagem e avaliação dos estudantes. Para Noble (2018), é incorreto concebê-la como neutra, seus resultados dependem da qualidade e da representatividade dos dados, bem como das escolhas metodológicas feitas por seus desenvolvedores. Assim, no campo educacional, decisões automatizadas, como avaliações ou recomendações de conteúdo, podem favorecer determinados grupos sociais em detrimento de outros, ampliando desigualdades já existentes. Dessa forma, a presença da IA nas escolas requer não apenas supervisão humana contínua, mas também estratégias explícitas de mitigação de vieses, garantindo inclusão e justiça pedagógica.

Essa problemática não se restringe apenas às avaliações, mas alcança o próprio processo de aprendizagem. Fernandes *et al.* (2024) ressaltam que, embora ofereça avanços importantes para a personalização do ensino, seu uso inadequado pode levar a práticas prejudiciais, como o incentivo ao plágio. O recurso indiscriminado



a ferramentas de geração automática de textos pode induzir estudantes a recorrerem ao “copiar e colar” sem análise crítica, comprometendo tanto a integridade acadêmica quanto o desenvolvimento de competências cognitivas fundamentais, como o pensamento crítico e a capacidade de síntese. A dependência de soluções rápidas, em detrimento de um aprendizado aprofundado, fragiliza a autonomia intelectual e empobrece a experiência formativa.

Nesse cenário, compete às instituições educacionais formular políticas explícitas de uso ético das tecnologias, promover processos de alfabetização digital crítica e incentivar práticas que fortaleçam a autoria, a reflexão acadêmica e a integridade intelectual dos estudantes.

Outro aspecto fundamental diz respeito à gestão dos dados educacionais. Boulay (2023) enfatiza que a adoção de sistemas digitais no contexto escolar exige atenção rigorosa à posse, ao uso e à proteção das informações dos estudantes. Para o autor, qualquer solução tecnológica aplicada ao ensino, à tutoria ou ao aconselhamento deve assumir um compromisso explícito com o cuidado integral do aluno, observando princípios de privacidade, equidade e diversidade. Nesse sentido, tais recursos não podem ser compreendidos apenas como mecanismos de automação; sua implementação requer políticas sólidas de governança de dados e supervisão docente contínua, a fim de evitar práticas discriminatórias e assegurar condições de justiça educacional.

Exemplos concretos demonstram os riscos associados a esses vieses. Guo *et al.* (2025) identificaram que estudantes para os quais o inglês não é língua materna são sistematicamente subavaliados por algoritmos de correção de textos, produzindo danos representacionais e alocativos que reforçam desigualdades (Kleinberg *et al.*, 2018). No mesmo sentido, Gándara *et al.* (2024) mostram que sistemas de predição de risco de evasão escolar, quando treinados em bases de dados desbalanceadas, classificam equivocadamente estudantes de minorias raciais ou socioeconômicas, negando-lhes apoio pedagógico essencial. Tais casos ilustram como os vieses algorítmicos não apenas distorcem avaliações, mas também perpetuam desigualdades educacionais historicamente enraizadas.

Diante desses riscos, emergem implicações éticas e pedagógicas inescapáveis. Amiri (2025) defende que sistemas de IA aplicados à educação sejam auditáveis, explicáveis e supervisionados por humanos, de modo a evitar decisões



opacas com impactos significativos sobre trajetórias estudantis. Do ponto de vista pedagógico, confiar exclusivamente em algoritmos compromete a função formativa da avaliação, reduz a experiência educacional a métricas quantitativas e invisibiliza dimensões sociais e subjetivas. Selwyn (2019) acrescenta que a suposta neutralidade algorítmica mascara escolhas políticas e pedagógicas que influenciam diretamente a forma como o aprendizado é reconhecido.

Essa crítica é aprofundada por Kim (2024), ao alertar para o risco de se considerar a IA como substituta integral do professor. Segundo o autor, tal concepção ignora a dimensão humana essencial do ensino, que envolve a escuta, o acolhimento e o diálogo reflexivo, elementos insubstituíveis da mediação docente. Assim, a mesma deve ser entendida como ferramenta complementar, capaz de apoiar processos de personalização, monitoramento e *feedback*, mas jamais substituir a presença humana, indispensável para interpretar contextos, compreender singularidades e fomentar a reflexão crítica. Nesse ponto, Selwyn (2022) e Boulay (2023) convergem ao afirmar que somente a integração equilibrada entre automação e supervisão humana pode assegurar uma educação inclusiva, justa e significativa.

Em síntese, a análise dos vieses algorítmicos na avaliação escolar revela que o problema não é apenas técnico, mas também ético e social. A questão central envolve não apenas como os algoritmos operam, mas, sobretudo, a quem eles servem e quem são os sujeitos prejudicados por suas decisões. Por isso, como defendem Nunes e Mercado (2025), torna-se urgente uma reformulação educacional sustentada por políticas públicas voltadas às populações menos favorecidas e capaz de acompanhar as transformações digitais contemporâneas. Somente assim a IA poderá se consolidar como uma aliada efetiva na construção de uma educação democrática e equitativa.

Estratégias para Mitigar Vieses Algorítmicos

A mitigação dos vieses algorítmicos em sistemas educacionais requer uma abordagem integrada, que articule vigilância técnica, transformação social e formação crítica. Estudos nacionais e internacionais indicam caminhos concretos para avançar nesse sentido, enfatizando que a tecnologia sozinha não garante justiça ou equidade na educação.



Um primeiro passo consiste na realização de auditorias regulares nos algoritmos, práticas capazes de identificar discrepâncias de desempenho entre diferentes grupos de estudantes. Barocas, Hardt e Narayanan (2023) destacam que tais auditorias funcionam como mecanismos de governança, revelando vieses implícitos que, de outra forma, permaneceriam invisíveis. Reforçando essa perspectiva, Heggler, Szmoski e Miquelin (2025) apontam que a verificação contínua é essencial desde a fase de codificação até os resultados automatizados, especialmente em contextos como o brasileiro, marcado por grande diversidade educacional e socioeconômica.

Avançando nessa discussão, a diversificação dos dados utilizados no treinamento de sistemas automatizados constitui uma estratégia essencial para reduzir desigualdades produzidas por modelos computacionais no contexto educacional. Essa diversificação significa incluir, de forma equilibrada, informações provenientes de escolas públicas e privadas, de áreas urbanas e rurais, além de incorporar as múltiplas variedades linguísticas e culturais presentes no país. Quando os dados contemplam apenas determinados grupos ou formas de expressão, o sistema tende a aprender padrões limitados, o que pode resultar em interpretações imprecisas e avaliações injustas.

Matsuoka (2023), ao analisar sistemas de correção automática de redações do Exame Nacional do Ensino Médio (ENEM), explica que um algoritmo só pode produzir resultados confiáveis quando é treinado com um conjunto de textos que reflita adequadamente a diversidade estrutural, temática e linguística dos estudantes brasileiros. Em outras palavras, quanto mais representativo for o *corpus*, menores são as chances de que o sistema penalize estudantes que escrevem com repertórios ou estruturas textuais diferentes daquelas mais frequentes nos dados originais.

No mesmo sentido, Silveira (2019) observa que a inclusão de bases de dados que expressem a pluralidade social não é apenas um aprimoramento técnico, mas uma necessidade ética e pedagógica. Sem essa pluralidade, torna-se provável que sistemas educacionais automatizados reproduzam desigualdades históricas, privilegiando determinados grupos enquanto marginalizam outros. Assim, a construção de tecnologias mais justas depende diretamente de escolhas conscientes sobre quais dados são coletados, como são organizados e de que modo representam a complexidade do contexto educacional brasileiro.



No campo das soluções técnicas, a aplicação de métodos de *fairness*, como *reweighting* e *balanced classes*, também se mostra eficaz para reduzir disparidades algorítmicas. Embora mais consolidadas em áreas como ciência da computação, tais abordagens já vêm sendo discutidas em sua adaptação para a educação. A noção de *fairness* (justiça algorítmica) refere-se a um conjunto de procedimentos que buscam minimizar vieses e garantir tratamento equitativo para diferentes grupos sociais, considerando aspectos como gênero, raça ou classe social (Kamiran; Calders, 2009; Mehrabi *et al.*, 2021).

Dentro desse escopo, o *reweighting* (reponderação de amostras) ajusta a importância de cada exemplo no conjunto de dados durante o treinamento do modelo, atribuindo pesos maiores a amostras de grupos sub-representados. Essa técnica assegura que o modelo aprenda a considerar todos os grupos de maneira equitativa, mesmo quando alguns são menos frequentes nos dados originais. Por exemplo, em um conjunto de dados de admissões escolares com predominância de estudantes urbanos, o *reweighting* aumenta o peso das amostras de estudantes rurais, garantindo que estes sejam avaliados com a mesma relevância. Por sua vez, as *balanced classes* (classes balanceadas) envolvem procedimentos para equilibrar a distribuição das classes no conjunto de dados. Quando uma classe está super-representada, o modelo tende a prever com maior precisão os casos dessa classe, negligenciando a minoritária e, conseqüentemente, gerando decisões injustas, como a negação de oportunidades a grupos menos favorecidos. Entre as estratégias de balanceamento, destacam-se o *oversampling* (sobreamostragem), que aumenta o número de amostras da classe minoritária por duplicação ou criação de novos exemplos, e o *undersampling* (subamostragem), que reduz a quantidade de amostras da classe majoritária, promovendo maior equilíbrio no conjunto de dados. Essas técnicas permitem que o modelo aprenda de forma mais justa, considerando adequadamente todas as classes (Kamiran; Calders, 2009; Mehrabi *et al.*, 2021). Nessa mesma linha, Vincent-Lancrin e Van Der Vlies (2020) reforçam que a incorporação de métricas de justiça e explicabilidade é condição essencial para assegurar legitimidade e aceitação social no uso da IA na educação.

Do ponto de vista pedagógico, a formação crítica de professores aparece como elemento indispensável para a mitigação de vieses. É necessário capacitar docentes para compreenderem os limites desses sistemas, interpretar seus resultados e intervir



pedagogicamente quando necessário. Vicari *et al.* (2023) defendem que a integração da IA na educação básica só se torna efetiva quando os professores são preparados para questionar, contextualizar e adaptar as tecnologias às necessidades de seus alunos. Nessa mesma direção, Russell (2021) destaca que a postura reflexiva dos docentes diante do seu uso é determinante para evitar que práticas avaliativas e pedagógicas se tornem automatizadas de forma acrítica.

Além disso, políticas públicas e marcos regulatórios cumprem papel estruturante nesse processo. A proposta da Lei de IA no Brasil (PL 2338/2023) e os debates internacionais sobre o "EU AI Act" (Lei da IA da União Europeia) reforçam a necessidade de transparência, direito à contestação e supervisão humana em sistemas de IA educacional (Floridi *et al.*, 2018; Brasil, 2023). Na mesma linha, Dourado (2019) argumenta que políticas educacionais devem priorizar a equidade, garantindo que a adoção de tecnologias não amplie desigualdades sociais e regionais. O Centro de Inovação para a Educação Brasileira (CIEB), acrescenta que a implementação na educação básica deve observar critérios de equidade, supervisão humana e adequação às diversidades regionais (CIEB, 2024).

A seguir, é sistematizado, com base nos argumentos aqui levantados nesta seção, um conjunto de estratégias voltadas à mitigação de vieses algorítmicos em sistemas educacionais, organizadas em diferentes dimensões de atuação (Quadro 1).

Quadro 1 - Resumo estratégias para mitigar vieses algorítmicos em sistemas educacionais

Dimensão	Ações Recomendadas	Referências
Vigilância técnica	<ul style="list-style-type: none"> Realizar auditorias regulares para identificar discrepâncias entre grupos de estudantes. Monitorar continuamente desde a fase de codificação até os resultados automatizados. 	Barocas; Hardt; Narayanan (2023); Heggler; Szmoski; Miquelin (2025)
Dados de treinamento	<ul style="list-style-type: none"> Diversificar bases de dados incluindo escolas públicas/privadas, contextos urbanos/rurais e variedades linguísticas regionais. Garantir que o corpus represente a heterogeneidade da produção textual dos estudantes. 	Matsuoka (2023); Silveira (2019)
Técnicas de <i>fairness</i>	<ul style="list-style-type: none"> Aplicar <i>reweighting</i> para atribuir maior peso a grupos sub-representados. 	Kamiran; Calders (2009); Mehrabi <i>et al.</i> (2021)

	<ul style="list-style-type: none"> • Utilizar <i>balanced classes</i> com estratégias de <i>oversampling</i> e <i>undersampling</i> para equilibrar distribuições. 	
Justiça e explicabilidade	<ul style="list-style-type: none"> • Incorporar métricas de justiça e de explicabilidade para fortalecer legitimidade e aceitação social dos sistemas. 	Vincent-Lancrin; Van Der Vlies (2020)
Formação docente	<ul style="list-style-type: none"> • Capacitar professores para compreender limites e potencialidades da IA. • Incentivar postura crítica e reflexiva frente ao uso pedagógico de sistemas automatizados. 	Vicari <i>et al.</i> (2023); Russell (2021)
Políticas públicas e regulação	<ul style="list-style-type: none"> • Garantir transparência, supervisão humana e direito à contestação nos sistemas de IA. • Estabelecer políticas que priorizem a equidade e respeitem diversidades regionais. • Implementar critérios claros para uso da IA na educação básica. 	Floridi <i>et al.</i> (2018); Brasil (2023); Dourado (2019); CIEB (2024)

Fonte: Elaborado pelos autores (2025)

No Quadro 1, o resumo das propostas procura articular aspectos técnicos, pedagógicos e regulatórios, destacando que a promoção da equidade no uso da IA na educação requer, simultaneamente, rigor metodológico no tratamento de dados, formação crítica dos professores e participação ativa de diversos atores sociais. A síntese indica caminhos para uma abordagem sociotécnica, na qual justiça, transparência e valorização da diversidade cultural e regional constituem princípios centrais no desenvolvimento e na aplicação desses sistemas.

Dessa perspectiva, a mitigação efetiva dos vieses exige uma abordagem sociotécnica e colaborativa, integrando desenvolvedores, gestores, educadores, pesquisadores e legisladores. Não basta otimizar algoritmos, é necessário contextualizá-los, auditá-los, interpretá-los e confrontar seus resultados com avaliações humanas e pedagógicas. Somente por meio dessa atuação conjunta será possível construir sistemas justos, contextualizados e comprometidos com os princípios da democracia e da justiça social.

Considerações finais

O presente estudo partiu do problema de investigar como os vieses algorítmicos se manifestam na avaliação escolar e quais são suas implicações éticas e pedagógicas. A análise teórico-reflexiva, fundamentada em revisão bibliográfica interpretativa, permitiu compreender que, embora a IA represente oportunidades significativas para a personalização do ensino e para a otimização do trabalho docente, sua utilização não está isenta de riscos. Entre eles, destacam-se a reprodução de desigualdades históricas, a redução da autonomia crítica dos sujeitos e a possibilidade de decisões automatizadas injustas.

Os principais resultados evidenciam que os vieses podem emergir tanto das bases de dados quanto do design dos algoritmos, impactando diretamente processos de avaliação, recomendação de conteúdos e monitoramento de desempenho. Isso revela que tais distorções não são apenas problemas técnicos, mas também sociais, culturais e éticos, exigindo uma abordagem que articule diferentes dimensões. Nesse sentido, o estudo destacou a importância de estratégias como auditorias regulares, diversificação de dados, aplicação de técnicas de *fairness*, formação crítica de professores e criação de marcos regulatórios que priorizem equidade, transparência e supervisão humana.

As contribuições deste trabalho situam-se, sobretudo, na problematização da ideia de neutralidade algorítmica e na defesa da construção de uma abordagem sociotécnica para a IA educacional, que reconheça o papel insubstituível do professor como mediador crítico e ético. Além disso, o estudo reforça que a democratização do acesso à tecnologia deve estar acompanhada de políticas públicas capazes de mitigar desigualdades regionais, socioeconômicas e culturais.

Entretanto, reconhece-se como limitação desta pesquisa o caráter eminentemente teórico-reflexivo, que não foi acompanhado de investigações empíricas em contextos educacionais concretos. Assim, recomenda-se que estudos futuros explorem a aplicação prática das estratégias aqui discutidas, por meio de experimentações em escolas, avaliações comparativas de sistemas de IA e análise da percepção de professores e estudantes quanto à utilização dessas tecnologias.

Em síntese, a consolidação de sistemas de IA justos e inclusivos na educação depende de uma atuação colaborativa entre desenvolvedores, gestores, educadores, pesquisadores e legisladores. Mais do que aperfeiçoar algoritmos, é necessário situá-



los em contextos pedagógicos e sociais, garantindo que estejam alinhados com os princípios da democracia, da justiça social e da valorização da diversidade cultural.

Referências

AMIRI, S. M. H. Digital Transformations in Education: Research Insights for 21st-Century Learning. **International Journal of Innovative Science, Engineering & Technology (IJISSET)**, [S.l.], v. 12, n. 03, p. 1-15, 2025. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5194886. Acesso em: 04 jun. 2025.

BAROCAS, S.; HARDT, M.; NARAYANAN, A. **Fairness and machine learning: Limitations and opportunities**. Cambridge: MIT press, 2023.

BOULAY, B. Inteligência artificial na educação e ética. **RE@ D-Revista de Educação a Distância e Elearning**, Lisboa, v. 6, n. 1, p. 1-17, 2023. Disponível em: https://revistas.rcaap.pt/lead_read/article/view/32242. Acesso em: 01 jun. 2025.

BRASIL. **Projeto de Lei nº 2338, de 2023**. Estabelece princípios, direitos e deveres para o uso de inteligência artificial no Brasil. Brasília: Congresso Nacional, 2023.

CASAGRANDE, R. **Inteligência Artificial e a Educação Além da Curva**. Curitiba: Edição do autor, 2019.

CIEB, Centro de Inovação para a Educação Brasileira. **Inteligência artificial na educação básica: novas aplicações e tendências para o futuro**. (CIEB Notas Técnicas, n. 21). São Paulo: CIEB, 2024.

DOURADO, L. F. Políticas e gestão da educação básica no Brasil: limites e perspectivas. **Educação & Sociedade**, Campinas, v. 40, p. 921-946, 2019. Disponível em: <https://www.scielo.br/j/es/a/w6QjW7pMDpzLrfRD5ZRkMWr/?format=html&lang=pt>. Acesso em: 01 jun. 2025.

FERNANDES, A. B. *et al.* A ética no uso de inteligência artificial na educação: implicações para professores e estudantes. **Revista Ibero-Americana de Humanidades, Ciências e Educação**, São Paulo, v. 10, n. 3, p. 346-361, 2024. Disponível em: <https://periodicorease.pro.br/rease/article/view/13056>. Acesso em: 01 jun. 2025.

FLORIDI, L. *et al.* AI4People - An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. **Minds and Machines**, [S.l.], v. 28, n. 4, p. 689-707, 2018. Disponível em: <https://link.springer.com/article/10.1007/S11023-018-9482-5>. Acesso em: 23 maio 2025.

GÁNDARA, D. *et al.* Inside the black box: detecting and mitigating algorithmic bias across racialized groups in college student-success prediction. **AERA Open**, [S.l.], v.

10, n. 1, p. 1-15, 2024. Disponível em:
<https://journals.sagepub.com/doi/full/10.1177/23328584241258741>. Acesso em: 22 maio 2025.

GUO, S. *et al.* Artificial Intelligence Bias on English Language Learners in Automatic Scoring. *In: International Conference on Artificial Intelligence in Education*, 26., 2025. Cham. **Anais [...]**, Cham: Springer Nature Switzerland, 2025. p. 268-275. Disponível em: https://link.springer.com/chapter/10.1007/978-3-031-98462-4_34. Acesso em: 02 jun. 2025.

HEGGLER, J.; SZMOSKI, R. M.; MIQUELIN, A. F. As dualidades entre o uso da inteligência artificial na educação e os riscos de vieses algorítmicos. **Educação & Sociedade**, Campinas, v. 46, p. 1-20, 2025. Disponível em:
<https://www.scielo.br/j/es/a/qrTryFvZR9Y9WsRpG5fWGHb/?format=html&lang=pt>. Acesso em: 04 abr. 2025.

HOLMES, W.; BIALIK, M.; FADEL, C. **Artificial intelligence in education promises and implications for teaching and learning**. Boston: Center for Curriculum Redesign, 2019.

KAMIRAN, F.; CALDERS, T. Data preprocessing techniques for classification without discrimination. **Knowledge and information systems**, [S.l.], v. 33, n. 1, p. 1-33, 2012. Disponível em: <https://link.springer.com/article/10.1007/s10115-011-0463-8>. Acesso em: 16 jun. 2025.

KIM, J. Leading teachers' perspective on teacher-AI collaboration in education. **Education and information technologies**, [S.l.], v. 29, n. 7, p. 8693-8724, 2024. Disponível em: <https://link.springer.com/article/10.1007/s10639-023-12109-5>. Acesso em: 03 maio 2025.

KLEINBERG, J. *et al.* Discrimination in the Age of Algorithms. **Journal of legal analysis**, Oxford, v. 10, p. 113-174, 2018. Disponível em:
<https://academic.oup.com/jla/article/doi/10.1093/jla/laz001/5476086?ref=akusion-ci-shi-dai-bizinesumedeia>. Acesso em: 02 jul. 2025.

LEE, M. K. Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. **Big data & society**, [S.l.], v. 5, n. 1, p. 1-16, 2018. Disponível em:
<https://journals.sagepub.com/doi/full/10.1177/2053951718756684>. Acesso em: 05 maio 2025.

LIMA, T. C. S.; MIOTO, R. C. T. Procedimentos metodológicos na construção do conhecimento científico: a pesquisa bibliográfica. **Revista katálysis**, Florianópolis, v. 10, p. 37-45, 2007. Disponível em:
<https://www.scielo.br/j/rk/a/HSF5Ns7dkTNjQVpRyvhc8RR/abstract/?lang=pt>. Acesso em: 02 abr. 2025.

LUCKIN, R. *et al.* **Intelligence unleashed: An argument for AI in education**. London: Pearson, 2016.

MATSUOKA, F. A. Automatic Essay Scoring in a Brazilian Scenario. **arXiv preprint**, Ithaca, p. 1-6, 2023. Disponível em: <https://arxiv.org/pdf/2401.00095>. Acesso em: 04 abr. 2025.

MEHRABI, N. *et al.* A survey on bias and fairness in machine learning. **ACM Computing Surveys (CSUR)**, [S.l.], v. 54, n. 6, p. 1-35, 2021. Disponível em: <https://dl.acm.org/doi/abs/10.1145/3457607>. Acesso em: 02 abr. 2025.

MENTA, E.; BRITO, G. S. O papel da Inteligência Artificial no Ensino Tecnológico: implicações emergentes. **Educitec - Revista de Estudos e Pesquisas sobre Ensino Tecnológico**, Manaus, v. 10, n. jan./dez., p. 1-22, 2024. Disponível em: <https://doi.org/10.31417/educitec.v10.2325>. Acesso em: 11 abr. 2025.

NOBLE, S. **Algorithms of oppression: how search engines reinforce racism**. New York: NYU Press, 2018.

NUNES, M. A. S.; MERCADO, L. P. L. Impacto da inteligência artificial na educação básica: equidade e desafios. **Revista Docência e Cibercultura**, Rio de Janeiro, v. 9, n. 1, p. 1-19, 2025. Disponível em: <https://www.e-publicacoes.uerj.br/re-doc/article/view/83188>. Acesso em: 02 jul. 2025.

RUSSELL, S. **Inteligência Artificial a nosso favor: como manter o controle sobre a tecnologia**. São Paulo: Companhia das Letras, 2021.

SELWYN, N. **Should robots replace teachers? AI and the future of education**. Cambridge: Polity Press, 2019.

SELWYN, N. **Education and technology: Key issues and debates**. 3. ed. London: Bloomsbury Academic, 2022.

SILVA, D. N.; COSTA, A. M. Technological Education and Innovation: perspectives for the Near Future. **Educitec - Revista de Estudos e Pesquisas sobre Ensino Tecnológico**, Manaus, v. 10, n. jan./dez., p. 1-8, 2024. Disponível em: <https://doi.org/10.31417/educitec.v10.2348>. Acesso em: 11 abr. 2025.

SILVEIRA, S. A. **Democracia e os códigos invisíveis: como os algoritmos estão modulando comportamentos e escolhas políticas**. São Paulo: Edições Sesc, 2019.

SOUSA, A. S.; OLIVEIRA, G. S.; ALVES, L. H. A pesquisa bibliográfica: princípios e fundamentos. **Cadernos da FUCAMP**, Monte Carmelo, v. 20, n. 43, p.64-83, 2021. Disponível em: <https://revistas.fucamp.edu.br/index.php/cadernos/article/view/2336>. Acesso em: 15 jun. 2025.

SURESH, H.; GUTTAG, J. V. A framework for understanding sources of harm throughout the machine learning life cycle. **arXiv preprint**, Ithaca, p. 1-8, v. 5, 1 dez. 2021. Disponível em: <https://arxiv.org/abs/1901.10002>. Acesso em: 15 jun. 2025.

TANG, L.; SU, Y. S. Ethical implications and principles of using artificial intelligence models in the classroom: a systematic literature review. **International Journal of Interactive Multimedia and Artificial Intelligence**, [S.l.], v. 8, n. 5, p. 25-36, 2024. Disponível em: <https://reunir.unir.net/handle/123456789/16210>. Acesso em: 03 jun. 2025.

VICARI, R. M. *et al.* **Inteligência artificial na educação básica**. São Paulo: Novatec, 2023.

VIEIRA, S. S. **Vieses algorítmicos são desafio para empresas na incorporação da IA**. TI Inside, 4 abr. 2025. Disponível em: <https://tiinside.com.br/04/04/2025/vieses-algoritmicos-sao-desafio-para-empresas-na-incorporacao-da-ia/>. Acesso em: 13 maio. 2025.

VILLANI, C. *et al.* **Donner un sens à l'intelligence artificielle**: pour une stratégie nationale et européenne. Rapport remis en mars 2018 à M. le Premier ministre (France). Paris: Vie-publique, 2018. Disponível em: <https://www.vie-publique.fr/rapport/37225-donner-un-sens-lintelligence-artificielle-pour-une-strategie-nation>. Acesso em: 03 jun. 2025.

VINCENT-LANCRIN, S.; VAN DER VLIES, R. Trustworthy artificial intelligence (AI) in education: Promises and challenges. **OECD education working papers**, Paris, n. 218, p. 1-17, 2020. Disponível em: <https://www.proquest.com/openview/6e6569f39d7c3b5240deac82103543a1/1?pq-origsite=gscholar&cbl=54479>. Acesso em: 03 jun. 2025.

ZAWACKI-RICHTER, O. *et al.* Systematic review of research on artificial intelligence applications in higher education – where are the educators? **International Journal of Educational Technology in Higher Education**, [S.l.], v. 16, n. 39, p. 1-27, 2019. Disponível em: <https://link.springer.com/article/10.1186/S41239-019-0171-0>. Acesso em: 05 maio 2025.

Recebido: 04/09/2025

Aprovado: 14/12/2025

Publicado: 26/02/2026

Como citar (ABNT): NASCIMENTO, M. S.; MARTINS, S. P.; MOSSETTE, A. S.; SANTOS, E. J. dos. Inteligência artificial e avaliação escolar: uma análise crítica dos vieses algorítmicos. **Educitec - Revista de Estudos e Pesquisas sobre Ensino Tecnológico**, Manaus, v. 12, e280326, 2026.

Contribuição de autoria:

Márcio Silveira Nascimento: Conceituação, investigação, curadoria de dados, metodologia e escrita (rascunho original).

Sidney Pires Martins: Escrita (revisão e edição).

Aline Santos Mossette: Escrita (revisão e edição).

Éber José dos Santos: Supervisão e escrita (revisão e edição).

Editor responsável: Iandra Maria Weirich da Silva Coelho

Direito autoral: Este artigo está licenciado sob os termos da Licença Creative Commons-Atribuição 4.0 Internacional

