DA NEUTRALIDADE ALGORÍTMICA AO JULGAMENTO PEDAGÓGICO: VIESES DA INTELIGÊNCIA ARTIFICIAL GENERATIVA E SEUS IMPACTOS NA JUSTIÇA AVALIATIVA NO ENSINO SUPERIOR

DE LA NEUTRALIDAD ALGORÍTMICA AL JUICIO PEDAGÓGICO: SESGOS DE LA INTELIGENCIA ARTIFICIAL GENERATIVA Y SUS IMPACTOS EN LA JUSTICIA EVALUATIVA EN LA EDUCACIÓN SUPERIOR

FROM ALGORITHMIC NEUTRALITY TO PEDAGOGICAL JUDGMENT: GENERATIVE ARTIFICIAL INTELLIGENCE BIASES AND THEIR IMPACTS ON ASSESSMENT FAIRNESS IN HIGHER EDUCATION

Viviane Cordeiro de QUEIROZ[1]
e-mail: vivicordeiroqueiroz35@gmail.com

Maria Hellena Ferreira BRASIL[2]
e-mail: hellenamhfb@gmail.com

Ana Cristina de Oliveira e SILVA[3]
e-mail: anacrisos@gmail.com

Como referenciar este artigo:

Queiroz, V., Brasil M. H., Silva, A. C. (2026). Da neutralidade algorítmica ao julgamento pedagógico: vieses da Inteligência Artificial Generativa e seus impactos na justiça avaliativa no Ensino Superior. *Rev. Int. Intelig. Artif. Tecnol. Educ. Form. Prof.*, *1*, e026002. https://doi.org/10.47519/IJAIET.v1i00.2

| **Submetido em**: 15/08/2026
| **Revisões requeridas em**: 09/09/2026
| **Aprovado em**: 23/10/2026
| **Publicado em**: 05/11/2026

Editores: Prof. Dr. José Anderson Santos Cruz
Profa. Dra. Beatriz da Silva Rosa Bonadiman

ARTIGO SUBMETIDO AO SISTEMA DE SIMILARIDADE

[1] Universidade Federal da Paraíba (UFPB), João Pessoa – Paraíba (PB) – Brasil. Doutoranda do Programa de Pós-Graduação em Enfermagem (PPGENF), vinculado ao Centro de Ciências da Saúde (CCS).
[2] Universidade Federal da Paraíba (UFPB), João Pessoa – Paraíba (PB) – Brasil. Doutoranda do Programa de Pós-Graduação em Enfermagem (PPGENF), vinculado ao Centro de Ciências da Saúde (CCS).
[3] Universidade Federal da Paraíba (UFPB), João Pessoa – Paraíba (PB) – Brasil. Professora permanente do Programa de Pós-Graduação em Enfermagem (PPGENF), vinculado ao Centro de Ciências da Saúde (CCS).

RESUMO: A incorporação da inteligência artificial generativa (IAG) à avaliação no ensino superior amplia possibilidades de automação, mas suscita preocupações com vieses algorítmicos e justiça avaliativa. Este estudo objetivou mapear e analisar evidências sobre vieses da IAG e sua influência no julgamento pedagógico docente, identificando implicações para a justiça avaliativa e a governança pedagógica. Realizou-se revisão de escopo, orientada pelo Joanna Briggs Institute (JBI) e pelo PRISMA-ScR, com buscas nas bases Scopus, Web of Science, ERIC, ACM Digital Library, Google Scholar e OpenAlex, complementadas por busca reversa de referências. Dos 220 estudos identificados, 13 integraram a síntese. Os resultados indicam que a IAG pode gerar respostas enviesadas e influenciar decisões docentes por ancoragem e viés de automação. Como contribuição conceitual, propõe-se o Framework de Governança Pedagógica da Inteligência Artificial (FGPIA), fundamentado em transparência, auditoria, justiça avaliativa, letramento em IA e responsabilidade, para orientar práticas avaliativas mais críticas, transparentes, responsáveis e equitativas.

PALAVRAS-CHAVE: Inteligência Artificial Generativa. Vieses algorítmicos. Justiça avaliativa. Julgamento pedagógico. Ensino Superior.

RESUMEN: *La incorporación de la inteligencia artificial generativa (IAG) en la evaluación en la educación superior amplía las posibilidades de automatización, pero plantea preocupaciones sobre los sesgos algorítmicos y la justicia evaluativa. Este estudio tuvo como objetivo identificar y analizar evidencias sobre los sesgos de la IAG y su influencia en el juicio pedagógico del profesorado, examinando sus implicaciones para la justicia evaluativa y la gobernanza pedagógica. Se realizó una revisión de alcance siguiendo la metodología del Joanna Briggs Institute (JBI) y las directrices PRISMA-ScR. Se efectuaron búsquedas en Scopus, Web of Science, ERIC, ACM Digital Library, Google Scholar y OpenAlex, complementadas con una búsqueda retrospectiva de referencias. De los 220 estudios identificados, 13 se incluyeron en la síntesis. Los resultados indican que la IAG puede generar respuestas sesgadas e influir en las decisiones docentes mediante el sesgo de anclaje y el sesgo de automatización. Como contribución conceptual, se propone el Marco de Gobernanza Pedagógica de la Inteligencia Artificial (MGPIA), basado en la transparencia, la auditoría, la justicia evaluativa, la alfabetización en IA y la responsabilidad, para orientar prácticas evaluativas más críticas, transparentes, responsables y equitativas.*

PALABRAS CLAVE: *Inteligencia Artificial Generativa. Sesgos algorítmicos. Justicia Evaluativa. Juicio pedagógico. Educación superior.*

ABSTRACT: *The integration of generative artificial intelligence (GenAI) into assessment in higher education expands opportunities for automation but raises concerns about algorithmic bias and fairness in assessment. This study aimed to map and analyze evidence on biases in GenAI and their influence on educators' pedagogical judgment, identifying implications for assessment fairness and pedagogical governance. A scoping review was conducted following the Joanna Briggs Institute (JBI) methodology and PRISMA-ScR guidelines. Searches were performed in Scopus, Web of Science, ERIC, ACM Digital Library, Google Scholar, and OpenAlex, supplemented by backward reference searching. Of the 220 studies identified, 13 were included in the synthesis. The findings indicate that GenAI can generate biased outputs and influence educators' decisions through anchoring and automation bias. As a conceptual contribution, the study proposes the Artificial Intelligence Pedagogical Governance Framework (AIPGF), grounded in transparency, auditing, assessment fairness, AI literacy, and accountability, to guide more critical, transparent, responsible, and equitable assessment practices.*

KEYWORDS: *Generative Artificial Intelligence. Algorithmic bias. Assessment fairness. Pedagogical judgment. Higher education.*

INTRODUÇÃO

A Inteligência Artificial Generativa (IAG) vem sendo incorporada às práticas docentes no ensino superior, incluindo a elaboração de questões, a construção de rubricas, a correção de produções acadêmicas e a geração de feedbacks. Essa adoção é impulsionada, sobretudo, pela promessa de eficiência, escalabilidade e apoio à tomada de decisão pedagógica (Zhao et al., 2024). Entretanto, a crescente utilização desses sistemas ocorre em meio à percepção de que os resultados produzidos por tecnologias baseadas em processamento estatístico seriam tecnicamente neutros e, portanto, potencialmente mais objetivos e justos do que os julgamentos exclusivamente humanos (Al Mohsen, 2025).

Essa percepção é problemática. Os modelos de IAG são desenvolvidos a partir de grandes conjuntos de dados que carregam marcas sociais, culturais, linguísticas e epistemológicas, enquanto os processos de curadoria, treinamento e alinhamento também envolvem escolhas humanas. Consequentemente, esses sistemas podem reproduzir ou amplificar assimetrias presentes nos dados e nas estruturas sociais que os constituem (Küçükuncular, 2026; Olaniyan et al., 2026). No campo educacional, essa questão adquire particular relevância, uma vez que as decisões avaliativas interferem diretamente no reconhecimento do desempenho, na trajetória acadêmica e nas oportunidades dos estudantes (Al Mohsen, 2025).

Embora a literatura sobre Inteligência Artificial na educação tenha avançado na identificação de vieses, riscos éticos e desafios de governança (Du et al., 2026), permanece menos explorada a dimensão especificamente pedagógica desse fenômeno, ou seja, como a mediação da IAG pode influenciar o próprio processo pelo qual o professor interpreta evidências, pondera informações e toma decisões sobre o desempenho discente. Assim, a lacuna não se restringe ao viés contido na saída algorítmica, mas envolve também os efeitos da interação entre essa saída e o julgamento humano. Diante disso, formulou-se a seguinte questão norteadora: como os vieses da Inteligência Artificial Generativa podem influenciar o julgamento pedagógico do professor e comprometer a justiça avaliativa no ensino superior? Objetivou-se mapear e analisar evidências sobre os vieses da Inteligência Artificial Generativa e sua influência sobre o julgamento pedagógico docente em processos avaliativos no ensino superior, identificando suas implicações para a justiça avaliativa e os princípios de governança pedagógica.

ARGUMENTAÇÃO TEÓRICA

A utilização da IAG na avaliação educacional abrange a geração de questões, a elaboração de rubricas, a produção de feedbacks, o apoio à correção e o suporte à tomada de decisão docente (Zhao et al., 2024; Selvam & González Vallejo, 2025). A rápida incorporação dessas tecnologias, por vezes anterior à elaboração de políticas institucionais suficientemente claras, amplia a possibilidade de usos acríticos e de deslocamento não deliberado de parte da autoridade avaliativa para os sistemas algorítmicos (Jiao et al., 2024).

A chamada neutralidade algorítmica corresponde à pressuposição de que sistemas baseados em procedimentos computacionais e cálculos estatísticos produziriam resultados isentos de valores e perspectivas. Essa concepção, entretanto, desconsidera que os algoritmos são desenvolvidos em contextos sociais e que tanto os dados quanto as escolhas envolvidas em seu treinamento podem incorporar assimetrias (Al Mohsen, 2025; Wang & Luo, 2026). Nos *datasets*, podem existir sub-representações linguísticas, culturais e epistemológicas; nos processos de treinamento e alinhamento, as escolhas humanas contribuem para moldar o comportamento dos modelos; e, diante da opacidade característica de sistemas complexos, torna-se difícil compreender integralmente os fatores que produziram determinada saída (Olaniyan et al., 2026).

Na avaliação educacional, tais características podem se manifestar de diferentes maneiras. Evidências apontam para vieses relacionados à raça, à etnia, ao gênero e ao idioma, além de dificuldades na consideração de contextos culturais e epistemológicos distintos (Wang & Luo, 2026; Yang et al., 2025; Liang et al., 2023; Mundim et al., 2025; Olaniyan et al., 2026). Esses fenômenos são particularmente relevantes porque uma avaliação justa não depende exclusivamente da precisão técnica de um sistema, mas também da equidade, da contextualização, da transparência e da possibilidade de responsabilização.

Nesse contexto, a justiça avaliativa pode ser compreendida como a garantia de que as decisões sobre o desempenho acadêmico sejam produzidas mediante processos equitativos, transparentes e passíveis de escrutínio, sem que características individuais ou contextuais dos estudantes sejam convertidas em desvantagens indevidas. Os modelos *human-in-the-loop* representam uma possibilidade de preservação da supervisão humana, desde que a participação do professor não seja meramente formal, mas envolva a capacidade real de questionar, revisar e rejeitar as recomendações produzidas pelo sistema (Selvam & González Vallejo, 2025).

CC BY

Assim, a questão central deixa de se limitar à análise de se determinada saída da IA é correta ou incorreta e passa a incluir a forma como a decisão avaliativa é construída quando o julgamento docente é mediado por uma recomendação algorítmica. Essa mudança de perspectiva fundamenta a discussão sobre governança pedagógica (Selvam & González Vallejo, 2025).

MÉTODO

Trata-se de uma revisão de escopo baseada na abordagem metodológica do Joanna Briggs Institute (JBI) e orientada pelo PRISMA-ScR (Aromataris et al., 2024; Joanna Briggs Institute, 2020; Tricco et al., 2018). O protocolo foi registrado no *Open Science Framework* e elaborado conforme as recomendações para protocolos de revisão.

Para a formulação da questão de pesquisa e a definição da estratégia de busca, utilizou-se a estrutura População, Conceito e Contexto (PCC). A população foi delimitada ao âmbito do ensino superior; o conceito contemplou a Inteligência Artificial Generativa, os vieses algorítmicos, o julgamento pedagógico e a justiça avaliativa; e o contexto correspondeu às instituições de ensino superior. A questão norteadora foi: como os vieses da Inteligência Artificial Generativa podem influenciar o julgamento pedagógico do professor e comprometer a justiça avaliativa no ensino superior?

As buscas foram realizadas entre maio e junho de 2026 nas bases Scopus (Elsevier), Web of Science, ERIC e ACM Digital Library e complementadas por buscas na literatura cinzenta, por meio do Google Scholar e do OpenAlex, e pela busca retrospectiva de referências (*backward citation searching*). Foram combinados descritores controlados e termos não controlados, incluindo “Generative Artificial Intelligence”, “Algorithmic Bias”, “Higher Education” e “Assessment”, adaptados à sintaxe específica de cada fonte de informação.

Foram identificados 220 estudos. Após a remoção de 33 duplicatas, 187 registros foram submetidos à triagem por título e resumo; 65 foram avaliados na íntegra e 13 estudos foram incluídos na síntese. Foram considerados elegíveis estudos empíricos e/ou teóricos relacionados à utilização de IAG em processos avaliativos no ensino superior, publicados entre 2023 e 2026, em português, inglês ou espanhol, e disponíveis na íntegra. Foram excluídos estudos exclusivamente relacionados à educação básica, sem possibilidade de transposição para o ensino superior, resumos sem texto completo e investigações sobre IA preditiva sem relação

com a IA generativa. A síntese dos achados foi realizada por meio de análise temática, com a organização das evidências em categorias relacionadas à neutralidade algorítmica, aos vieses dos modelos, à influência da IAG sobre o julgamento docente e à justiça avaliativa. A proposição do *Framework* de Governança Pedagógica da Inteligência Artificial (FGPIA) resultou da interpretação integrada desses achados e não constitui um modelo empiricamente validado.

As ferramentas de IAG foram utilizadas exclusivamente como apoio técnico, em conformidade com a Portaria nº 2.664/2026 do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). O OpenEvidence apoiou a sumarização inicial dos achados, cujo conteúdo foi posteriormente conferido pelos autores. O SciSpace apoiou a organização das categorias temáticas. O QuillBot foi utilizado exclusivamente para a revisão ortográfica e a melhoria da fluidez textual. A seleção, a interpretação, a análise crítica e a redação final são de responsabilidade exclusiva dos autores.

RESULTADOS

A busca identificou 220 estudos, dos quais 13 foram incluídos na síntese após as etapas de identificação, remoção de duplicatas, triagem e avaliação de elegibilidade. A análise temática permitiu identificar quatro categorias principais, que expressam os eixos recorrentes nas evidências examinadas.

Categoria 1: Neutralidade Algorítmica

Os estudos analisados questionam a concepção de neutralidade técnica dos algoritmos, reconhecendo os sistemas de IA como produtos sociotécnicos influenciados pelos dados, pelas escolhas de desenvolvimento e pelos contextos em que são utilizados (Wang & Luo, 2026; Jiao et al., 2024; Ragolane et al., 2024). Os achados indicam que as assimetrias presentes nos dados e nos processos de desenvolvimento podem repercutir nas respostas produzidas pelos modelos, enquanto a chamada “insensibilidade contextual” pode limitar a adequação das avaliações automatizadas a diferentes realidades educacionais (Huang, 2025).

Categoria 2: Vieses dos Modelos de IA Generativa

Foram identificadas evidências de vieses relacionados à raça, à etnia, ao gênero, ao idioma, à cultura e à epistemologia. Os estudos apontam dificuldades dos modelos em reconhecer adequadamente características linguísticas e demográficas, bem como diferenças culturais e epistemológicas que podem interferir na avaliação de produções acadêmicas (Wang & Luo, 2026; Yang et al., 2025; Mundim et al., 2025; Olaniyan & Martins; Al Maqrashi, 2026). No campo linguístico, os detectores de texto gerado por IA podem apresentar viés contra produções de autores que não são falantes nativos de inglês, o que levanta preocupações quanto à equidade quando esses detectores são utilizados em contextos educacionais (Liang et al., 2023).

Categoria 3: Influência da IA sobre o Julgamento Docente

Esta categoria apresentou menor desenvolvimento no corpus analisado, evidenciando uma lacuna na produção científica diretamente voltada à interação entre recomendações algorítmicas e julgamento pedagógico. Estudos sobre a percepção de professores universitários indicam que ferramentas de IAG vêm sendo consideradas recursos de apoio à avaliação, especialmente em relação à eficiência e à redução da carga de trabalho docente (Khlaif et al., 2024).

Uma evidência experimental envolvendo 1.339 professores mostrou que os docentes corrigiram menos erros de pontuação quando a avaliação incorreta era atribuída a um sistema de IA do que quando era atribuída a um avaliador humano, sugerindo a presença de efeitos de automação sobre a revisão das decisões (Goulas et al., 2026). Outros achados demonstram que avisos algorítmicos e indicadores visuais podem modificar sistematicamente a avaliação docente de produções textuais (Du et al., 2026). Entretanto, a expertise humana permanece relevante para a interpretação de nuances contextuais e pedagógicas que não são necessariamente capturadas pelos sistemas automatizados (Ragolane et al., 2024; Huang et al., 2023).

Categoria 4: Justiça Avaliativa

Os estudos relacionam a justiça avaliativa à necessidade de assegurar que estudantes de diferentes origens não sejam prejudicados por vieses algorítmicos, desigualdades de acesso ou limitações dos sistemas utilizados na avaliação (Karunaratne et al., 2025). A supervisão humana aparece como elemento central, especialmente em modelos *human-in-the-loop*, nos quais a decisão final não é integralmente delegada ao sistema (Selvam & González Vallejo, 2025). Também são apontadas estratégias de redesenho das avaliações como forma de preservar a integridade acadêmica e os direitos dos estudantes diante da expansão da IAG (Lye & Lim, 2024).

ANÁLISE E DISCUSSÃO DOS DADOS

Os achados convergem para a rejeição da ideia de que a utilização da IAG na avaliação seja uma intervenção exclusivamente técnica. Os vieses identificados nos modelos revelam que raça, gênero, idioma, cultura e epistemologia podem interferir na forma como as produções acadêmicas são processadas e classificadas (Wang & Luo, 2026; Yang et al., 2025; Liang et al., 2023; Olaniyan et al., 2026). Entretanto, o principal achado interpretativo desta revisão é que o problema da justiça avaliativa não se encerra na saída produzida pelo sistema.

A mediação algorítmica introduz uma segunda dimensão: a possibilidade de a recomendação da IA influenciar o próprio julgamento do professor. Nesse sentido, identificam-se dois mecanismos com sustentação empírica mais direta no corpus: a ancoragem algorítmica e o viés de automação.

A ancoragem algorítmica ocorre quando uma recomendação produzida pela IA passa a constituir o ponto de referência inicial para a decisão docente. Em vez de construir sua avaliação a partir da análise direta da produção discente e utilizar a IA como recurso complementar, o professor pode tomar a saída algorítmica como referência para seu julgamento posterior. O experimento realizado com 1.339 professores, no qual os participantes corrigiram menos erros quando estes eram atribuídos à IA do que quando eram atribuídos a um avaliador humano, oferece evidência empírica relevante para esse fenômeno (Goulas et al., 2026).

Esse achado dialoga com o conceito clássico de ancoragem, segundo o qual os julgamentos podem ser influenciados por informações inicialmente apresentadas como

referência (Tversky & Kahneman, 1974). Evidências provenientes de outros campos, como a patologia computacional, também identificam efeitos de ancoragem e automação em decisões profissionais assistidas por sistemas computacionais (Rosbach et al., 2026). Contudo, a transposição desses achados para a avaliação educacional deve ser tratada como hipótese de investigação, e não como evidência causal direta no contexto pedagógico.

O segundo mecanismo é o viés de automação. Os estudos analisados indicam que a presença de alertas algorítmicos e indicadores visuais de risco pode alterar a forma como os professores avaliam produções acadêmicas, sugerindo que a recomendação automatizada não é cognitivamente neutra para o usuário (Du et al., 2026). Assim, mesmo quando o professor permanece formalmente responsável pela decisão, a existência de uma recomendação algorítmica pode modificar o grau de escrutínio dedicado à evidência original.

A partir desses achados, propõe-se ainda um terceiro movimento, de natureza teórica: a reconfiguração do repertório pedagógico pela padronização. Ao delegar etapas da avaliação a sistemas generativos, o docente pode passar a operar dentro das categorias, dos critérios e das formas de representação privilegiadas pelo sistema. Essa possibilidade encontra respaldo no debate crítico sobre a reorganização do trabalho docente mediado por tecnologias de IA (Küçükuncular, 2026), mas ainda carece de investigação empírica específica no contexto do julgamento pedagógico. Por essa razão, essa possibilidade é apresentada neste estudo como hipótese para pesquisas futuras.

A integração desses elementos desloca o eixo da justiça avaliativa. A pergunta não deve ser apenas “a nota produzida ou sugerida pela IA é justa?”, mas também “o processo pelo qual o professor chegou à decisão preservou sua autonomia crítica e a equidade entre os estudantes?”. A justiça passa, portanto, a envolver tanto a qualidade da saída algorítmica quanto a integridade do processo decisório.

Uma revisão sistemática recente evidencia a fragmentação da literatura sobre justiça, transparência, implicações pedagógicas e governança na avaliação mediada por IA, propondo a integração entre essas dimensões (Alfaleh, 2026). Em diálogo com essa perspectiva, apresenta-se a proposição do Framework de Governança Pedagógica da Inteligência Artificial (FGPIA). O framework não é apresentado como um instrumento validado, mas como uma contribuição conceitual destinada a transformar os achados da revisão em princípios operacionais para orientar o uso responsável da IAG em processos avaliativos.

Quadro 1 – Framework de Governança Pedagógica da IA (FGPIA)

Dimensão	Pergunta norteadora
Transparência	Como a IA chegou a essa resposta?
Auditoria	O professor verificou criticamente a saída da IA?
Justiça Avaliativa	A decisão preserva equidade entre os estudantes?
Letramento em IA	O docente compreende os vieses do sistema?
Responsabilidade	Quem assume a decisão final e responde por ela?

Fonte: Elaborado pelos autores, 2026.

As cinco dimensões articulam o problema identificado às respectivas respostas de governança. A transparência enfrenta a opacidade do sistema; a auditoria responde ao risco de aceitação acrítica das recomendações; a justiça avaliativa estabelece a equidade como critério explícito; o letramento em IA amplia a capacidade docente de reconhecer limitações e vieses; e a responsabilidade preserva a centralidade humana na decisão final.

O FGPIA, portanto, desloca a governança da IA de um conjunto genérico de recomendações técnicas para uma prática situada no próprio processo avaliativo. Sua finalidade não é impedir o uso da IAG, mas estabelecer condições para que sua utilização não substitua o julgamento pedagógico nem transforme a eficiência tecnológica em um critério superior à justiça educacional.

Como limitações do estudo, destacam-se a heterogeneidade dos estudos incluídos, o recorte temporal adotado e a própria natureza da revisão de escopo, que permite mapear evidências, mas não estabelecer relações causais. Em diálogo com modelos contemporâneos de governança da IA na avaliação, o FGPIA concentra-se especificamente no ponto de interação entre a recomendação algorítmica e o julgamento docente e ainda não foi submetido à validação empírica.

CONSIDERAÇÕES FINAIS

Esta revisão de escopo permitiu mapear evidências sobre os vieses da Inteligência Artificial Generativa e suas implicações para o julgamento pedagógico e a justiça avaliativa no ensino superior. Os estudos analisados evidenciam vieses relacionados à raça, à etnia, ao gênero, ao idioma, à cultura e à epistemologia, além de indicarem que as recomendações algorítmicas podem influenciar os processos de julgamento docente por meio de mecanismos associados à ancoragem e ao viés de automação. Os achados permitem sustentar que o desafio da IAG na avaliação educacional ultrapassa a questão da precisão da resposta produzida pelo

sistema. A justiça avaliativa depende também da preservação da autonomia crítica do professor, da possibilidade de auditoria das saídas, da compreensão dos limites da tecnologia e da responsabilização humana pela decisão final.

Como contribuição conceitual, propôs-se o FGPIA, estruturado em cinco dimensões: transparência, auditoria, justiça avaliativa, letramento em IA e responsabilidade. O framework pode subsidiar protocolos institucionais de utilização da IAG, processos de formação docente e mecanismos de revisão das decisões avaliativas mediadas por tecnologias generativas.

Recomenda-se, portanto, a realização de estudos com docentes do ensino superior para investigar a ocorrência de ancoragem e viés de automação em situações reais de avaliação, bem como a realização de estudos metodológicos destinados à validação das cinco dimensões do framework e ao desenvolvimento de instrumentos formativos para sua operacionalização.

REFERÊNCIAS

Alfaleh, M. (2026). Sustainable AI-driven assessment in higher education: A systematic review of fairness, transparency, pedagogical innovation, and governance. *Sustainability, 18*(2), Article 785. https://doi.org/10.3390/su18020785

Al Mohsen, F. S. (2025). Ethical integration of generative AI in higher education assessments. In A. ElSayary (Ed.), *Prompt engineering and generative AI applications for teaching and learning* (pp. 275–288). IGI Global. https://doi.org/10.4018/979-8-3693-7332-3.ch017

Aromataris, E., Lockwood, C., Porritt, K., Pilla, B., & Jordan, Z. (Eds.). (2024). *JBI manual for evidence synthesis*. JBI. https://doi.org/10.46658/JBIMES-24-01

Du, P., Liu, T., & Xian, X. (2026). Automation bias in teachers' evaluation of student writing: Effects of algorithmic warnings and visual risk cues in AI detection reports. *Frontiers in Psychology, 17*, Article 1889402. https://doi.org/10.3389/fpsyg.2026.1889402

Goulas, S., Megalokonomou, R., & Sotirakopoulos, P. (2026). Why do experts miss AI's errors? Evidence from a randomized labeling experiment. *PNAS Nexus, 5*(6), Article pgag146. https://doi.org/10.1093/pnasnexus/pgag146

Huang, S. (2025). AI in education: A critical review of applications, bias, and equity challenges. *Advances in Engineering Technology Research, 15*(1), Article 1679. https://doi.org/10.56028/aetr.15.1.1679.2025

Jiao, J., Afroogh, S., Chen, K., Atkinson, D., & Dhurandhar, A. (2024). *The global landscape of academic guidelines for generative AI and large language models* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2406.18842

Joanna Briggs Institute. (2020). *Template for scoping reviews protocols* [Template]. https://jbi.global/scoping-review-network/resources

Karunaratne, T., Aghaee, N., & Ferati, M. (2025). Harnessing the generative AI wave towards fair and diverse higher education assessments: A comprehensive analysis through an innovative lens of students. *European Conference on e-Learning, 24*(1), 476–484. https://doi.org/10.34190/ecel.24.1.4259

Khlaif, Z. N., Ayyoub, A., Hamamra, B., Bensalem, E., Mitwally, M. A. A., Ayyoub, A., Hattab, M. K., & Shadid, F. (2024). University teachers' views on the adoption and integration of generative AI tools for student assessment in higher education. *Education Sciences, 14*(10), Article 1090. https://doi.org/10.3390/educsci14101090

Küçükuncular, A. (2026). Dehumanising education: AI and the capitalist capture of teaching. *AI and Ethics, 6*, Article 123. https://doi.org/10.1007/s43681-026-01006-5

Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. *Patterns, 4*(7), Article 100779. https://doi.org/10.1016/j.patter.2023.100779

Lye, C. Y., & Lim, L. (2024). Generative artificial intelligence in tertiary education: Assessment redesign principles and considerations. *Education Sciences, 14*(6), Article 569. https://doi.org/10.3390/educsci14060569

Mundim, R. R., Dionisio, D. L., Mena, D. da S., Oliveira, D., & Victer, E. das F. (2025). Evaluation of open-ended responses in education with the help of ChatGPT: Challenges in implementation. *Revista de Gestão Social e Ambiental, 19*(9), Article e013330. https://doi.org/10.24857/rgsa.v19n9-044

Olaniyan, Y. D., Martins, M. O., & Al Maqrashi, R. H. (2026). Generative artificial intelligence and epistemic (in)justice: Perspectives from higher education students in the Global North and South. *Frontiers in Human Dynamics, 8*, Article 1790324. https://doi.org/10.3389/fhumd.2026.1790324

Ragolane, M., Patel, S., & Salikram, P. (2024). AI versus human graders: Assessing the role of large language models in higher education. *Asian Journal of Education and Social Studies, 50*(10), 244–263. https://doi.org/10.9734/ajess/2024/v50i101616

Rosbach, E., Ammeling, J., Ganz, J., Bertram, C. A., Conrad, T., Riener, A., & Aubreville, M. (2026). Stuck on suggestions: Automation bias, the anchoring effect, and the factors that shape them in computational pathology. *Machine Learning for Biomedical Imaging, 2026*, 126–147. https://doi.org/10.59275/j.melba.2026-87b1

Selvam, M., & González Vallejo, R. (2025). Human-in-the-loop models for ethical AI grading: Combining AI speed with human ethical oversight. *EthAIca, 4*, Article 413. https://doi.org/10.56294/ai2025413

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., . . . Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine, 169*(7), 467–473. https://doi.org/10.7326/M18-0850

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Wang, X., & Luo, G. (2026). *Evaluating racial bias in LLM reasoning: Implications for equitable AI use in education* [Preprint]. OSF Preprints. https://doi.org/10.35542/osf.io/ynt3h_v1

Yang, K., Raković, M., Gašević, D., & Chen, G. (2025). *Does the prompt-based large language model recognize students' demographics and introduce bias in essay scoring?* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2504.21330

Zhao, J., Chapman, E., & Sabet, P. G. P. (2024). Generative AI and educational assessments: A systematic review. *Education Research and Perspectives, 51*, 124–155. https://doi.org/10.70953/ERPv51.2412006

CRediT Author Statement

- **Reconhecimentos**: Não.
- **Financiamento**: Não houve financiamento.
- **Conflitos de interesse**: Não houve conflitos de interesse.
- **Aprovação ética**: Não precisou de aprovação ética.
- **Disponibilidade de dados e material**: Disponível nas referências.
- **Contribuições dos autores**: Todas as autoras contribuíram para a elaboração do trabalho.

Processamento e editoração: Editora Ibero-Americana de Educação
Revisão, formatação, normalização e tradução