

## Uso de IAs Generativas na Avaliação Automatizada da Aprendizagem: Um Mapeamento Sistemático da Literatura

Otávio Lube dos Santos, Universidade Federal do Espírito Santo, olube@inf.ufes.br  
Davidson Cury, Universidade Federal do Espírito Santo, dede@inf.ufes.br

**Resumo:** Este estudo realiza um Mapeamento Sistemático da Literatura (MSL) sobre o uso de IAs generativas na avaliação automatizada da aprendizagem. Foram analisados 40 artigos publicados entre janeiro de 2019 e setembro de 2024, selecionados de bases como Scopus, Web of Science e IEEE Xplore. A pesquisa identificou tendências, metodologias e resultados no uso de modelos de linguagem, como GPT-3 e GPT-4, aplicados em avaliações automáticas, principalmente em disciplinas de programação e línguas. Os resultados mostram que IAs generativas são eficazes para correção automatizada e feedback personalizado, muitas vezes combinadas com métodos convencionais de avaliação, garantindo maior confiabilidade. No entanto, desafios relacionados à precisão, aceitação por educadores e questões éticas, como privacidade e equidade, permanecem. O estudo conclui que, apesar do potencial significativo, a adoção em larga escala depende da superação desses desafios e da preparação adequada das instituições educacionais.

**Palavras-chave:** Inteligência Artificial Generativa, Avaliação Automatizada, Aprendizagem, Mapeamento Sistemático, Educação

### The Use of Generative AIs in Automated Learning Assessment: A Systematic Literature Mapping

**Abstract:** This study conducts a Systematic Literature Mapping (SLM) on the use of generative AIs in automated learning assessment. A total of 40 articles published between January 2019 and September 2024 were analyzed, selected from databases such as Scopus, Web of Science, and IEEE Xplore. The research identified trends, methodologies, and outcomes regarding the use of language models like GPT-3 and GPT-4, applied in automated assessments, particularly in programming and language disciplines. The results show that generative AIs are effective for automated grading and personalized feedback, often combined with conventional assessment methods, ensuring greater reliability. However, challenges related to accuracy, educator acceptance, and ethical issues such as privacy and equity remain. The study concludes that, despite the significant potential, large-scale adoption depends on overcoming these challenges and adequately preparing educational institutions.

**Keywords:** Generative Artificial Intelligence, Automated Assessment, Learning, Systematic Mapping, Education.

#### 1. Introdução

Nos últimos anos, a aplicação de Inteligência Artificial (IA) no campo da educação tem se expandido rapidamente, com destaque para as IAs generativas, que são capazes de criar textos, resolver problemas complexos e gerar conteúdos personalizados para os alunos. Essas tecnologias têm o potencial de transformar a forma como a avaliação educacional é realizada, especialmente em contextos de aprendizagem automatizada (FRANKFORD *et al.*, 2024; MESSER *et al.*, 2024). Tradicionalmente, a avaliação da aprendizagem tem sido limitada a métodos somativos e formativos, que, embora essenciais, enfrentam desafios relacionados à escalabilidade, personalização e entrega de feedback imediato (NICKL *et al.*, 2024; CLOUDE *et al.*, 2024).

Com o avanço de modelos de linguagem como o GPT (Generative Pre-trained Transformer), novas oportunidades surgem para automatizar processos de avaliação, permitindo uma análise mais detalhada do progresso dos alunos e a entrega de feedback em tempo real (LIANG e WU, 2024). Tais modelos possibilitam não apenas a correção automatizada de respostas, mas também a personalização do feedback, o que pode melhorar o desempenho dos alunos e desenvolver habilidades como pensamento crítico e criativo (LIANG e WU, 2024). Além disso, o uso de técnicas como *Learning Analytics* e Mineração de Dados Educacionais já trouxe avanços importantes no monitoramento do aprendizado, mas as IAs generativas têm o potencial de ir além, promovendo uma experiência de aprendizado ainda mais personalizada.

No entanto, apesar do potencial dessas tecnologias, a literatura sobre o uso de IAs generativas na avaliação automatizada da aprendizagem ainda carece de uma visão abrangente que mapeie o estado atual da pesquisa, identifique lacunas e direcione futuras investigações. Há também desafios éticos e técnicos, como a privacidade dos dados e a equidade no tratamento dos alunos, que precisam ser considerados.

Este trabalho tem como objetivo realizar um Mapeamento Sistemático da Literatura (MSL) sobre o uso de IAs generativas na avaliação automatizada da aprendizagem. O MSL se diferencia da revisão sistemática ao focar em mapear e categorizar a pesquisa existente, fornecendo uma visão geral das abordagens, ferramentas e métodos já empregados. Através desse mapeamento, buscamos identificar tendências, lacunas de pesquisa e oportunidades futuras no uso de IAs generativas na avaliação educacional (GÖRER e AYDEMIR, 2024; HICKMAN e BELL, 2024). Este estudo pretende apoiar pesquisadores e profissionais da educação na compreensão do estado da arte e na tomada de decisões fundamentadas sobre a adoção dessas tecnologias no ambiente educacional.

O trabalho está estruturado com este capítulo introdutório. Segue com a apresentação da metodologia no capítulo 2. Depois da análise e discussão dos resultados no capítulo 3 e das conclusões no capítulo 4. No capítulo 5 são apresentadas as referências utilizadas no estudo.

## **2. Metodologia**

Este trabalho apresenta um Mapeamento Sistemático e Literatura (MSL), cujo objetivo é fornecer uma visão geral de um campo de pesquisa amplo, categorizando e mapeando publicações realizadas e identificando áreas mais e menos exploradas (PETERSEN *et al.*, 2008).

Foram pesquisados trabalhos publicados entre jan/2019 e set/2024 a partir de buscas nas seguintes bases internacionais, acessadas por meio do Portal de Periódicos da CAPES: Scopus, Springer, Science Direct, Association for Computing Machinery Digital Library (ACM DL), Web of Science, Taylor & Francis Online e IEEE Xplore.

O período foi definido com o objetivo de identificar o crescimento da inserção de métodos de Inteligência Artificial Generativa em avaliações de estudantes em ambientes educacionais formais.

### **2.1. Questões de Pesquisa**

Para alcançar o objetivo principal deste trabalho foi definida uma questão principal (QP) a ser respondida: como usar a Inteligência Artificial para avaliação automatizada da aprendizagem?

Para apoiar a investigação, foram ainda elaboradas algumas questões secundárias, permitindo uma compreensão mais aprofundada dos resultados encontrados.

- QP1 - Quais os tipos de estudos e abordagens utilizadas?
- QP2 - Quais as metodologias de avaliação da aprendizagem utilizadas?
- QP3 - Quais as técnicas de Inteligência Artificial utilizadas?
- QP4 - Foi utilizando algum outro método convencional de avaliação?

## 2.2. Estratégia de Busca

A estratégia de busca incluiu as seguintes bases de dados: Scopus, Science Direct, Association for Computing Machinery Digital Library (ACM DL), Web of Science, Taylor & Francis Online e IEEE Xplore. As bases foram selecionadas considerando sua relevância para a área de computação e com a temática do trabalho. Tais periódicos foram acessados a partir do Portal de Periódicos da CAPES, que permitiu acesso a grande parte dos artigos aqui sintetizados.

Os mecanismos de busca *online* possibilitaram o uso de filtros por ano, palavras-chave e campos personalizados de pesquisa. Foi então construída a seguinte *string* de busca, observando-se as questões de pesquisa de interesse, a partir de palavras-chave.

**Quadro 1 - String de busca utilizada**

<p>(“Generative AI” OR “GenAI” OR “LLM” OR “ChatGPT” OR “OpenAI” OR  “Copilot” OR “Gemini” OR “CODEX” OR “GPT”)  AND  (“Automated Assessment” OR “Automated Feedback” OR “Learning  Assessment” OR “Student Evaluation” OR “AI-based Assessment” OR  “Assessment Systems”)</p>
--

## 2.3. Critérios de Inclusão e Exclusão

Na sequência, foram definidos critérios de inclusão e exclusão para filtrar os resultados com maior alinhamento à proposta do trabalho.

**Quadro 2 - Critérios de Inclusão**

Código	Critério de Inclusão
CI2	Estudos em inglês
CI4	Estudos primários
CI1	Artigos completos publicados entre jan/2019 e set/2024
CI5	Estudos devem ser aplicados no ensino médio ou superior
CI3	Estudos que abordem IA na avaliação automatizada da aprendizagem dos alunos

**Quadro 3 - Critérios de Exclusão**

Código	Critério de Exclusão
CE1	Artigo duplicado
CE2	Artigo secundário de Revisão Sistemática de Literatura
CE3	Estudos aplicados fora do contexto do ensino superior ou médio
CE4	Estudos sem utilização de IA
CE5	Estudo publicado fora de artigo completo

## 2.4. Seleção de Estudos

As buscas nas bases de dados deram origem a um total de 341 trabalhos potenciais, conforme pode ser observado na tabela 1.

**Tabela 1 - Avaliação de Trabalhos**

<b>Critério</b>	<b>Artigos Incluídos</b>	<b>Artigos Excluídos</b>
Scopus	4	26
Science Direct	15	196
ACM DL	8	46
Web of Science	3	16
Taylor & Francis Online	6	15
IEEE Xplore	2	4

O processo de análise e extração de dados respeitou o modelo Prisma, conforme (PAGE *et al.*, 2021), permitindo gerar o diagrama de fluxo da Figura 1, o qual ilustra o processo de avaliação dos trabalhos.

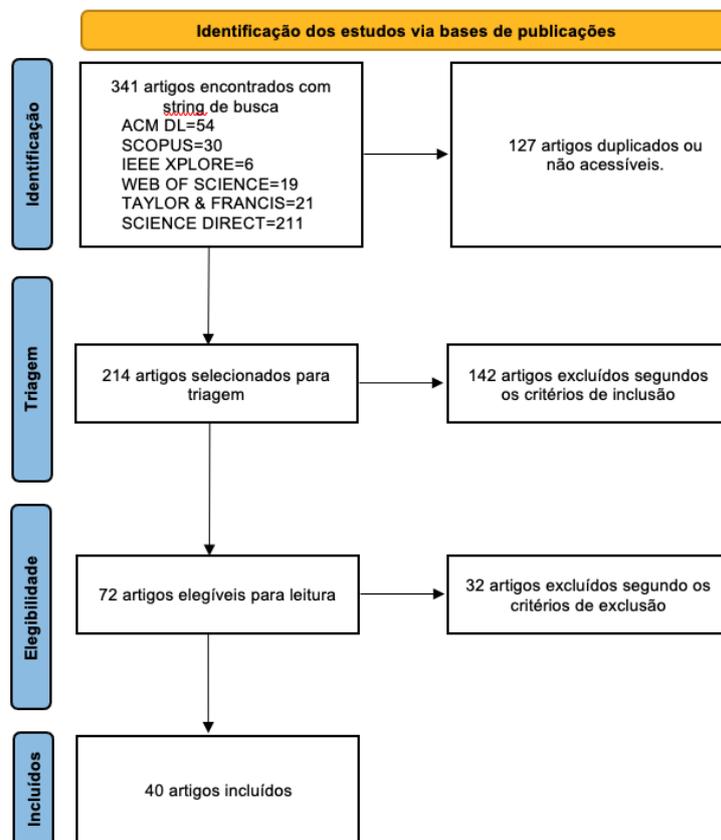


Figura 1. Diagrama de Fluxo - Prisma

### 3. Resultados e Discussão

Neste capítulo, são apresentados os principais resultados obtidos a partir da análise dos estudos incluídos no Mapeamento Sistemático da Literatura (MSL). O objetivo desta seção é examinar os dados coletados em relação às perguntas de pesquisa formuladas, permitindo uma compreensão mais aprofundada das tendências observadas no uso de IAs generativas na avaliação automatizada da aprendizagem.

A Figura 2 apresenta a distribuição geográfica das publicações incluídas no estudo. Observa-se que a maioria dos estudos está concentrada em países como Estados Unidos, China, Alemanha e Austrália, refletindo o avanço dessas nações em tecnologias educacionais baseadas em IA. Outros países, como Reino Unido, França,

Turquia e Irlanda, também aparecem com contribuições significativas, o que indica o interesse global no tema, embora haja uma concentração em certas regiões. Essa distribuição geográfica reforça a liderança de países desenvolvidos na pesquisa sobre IA aplicada à educação, mas também revela a necessidade de uma maior diversificação nas contribuições globais.

Quanto à evolução temporal das publicações, conforme ilustrado na Figura 3, verifica-se um aumento expressivo no número de estudos publicados nos últimos anos. Em 2023, o número de publicações ainda era modesto, com apenas algumas contribuições, mas em 2024 houve um salto significativo no volume de pesquisas, com mais de 30 estudos identificados. Esse crescimento sugere um aumento no interesse acadêmico e nas aplicações práticas das IAs generativas na educação, possivelmente impulsionado pelos avanços tecnológicos recentes, como o uso de modelos de linguagem como o GPT-3 e GPT-4.

Os resultados também são discutidos em consonância com a literatura atual, destacando-se pontos de convergência e divergência, bem como as limitações encontradas em alguns dos estudos analisados. Em particular, observa-se que, embora os avanços sejam promissores, desafios técnicos e éticos, como a privacidade de dados e a implementação em larga escala, ainda precisam ser amplamente discutidos e solucionados.

Por fim, são exploradas as implicações dos achados para o futuro da avaliação educacional e o impacto potencial das IAs generativas no ensino e na aprendizagem, considerando aspectos éticos e técnicos relevantes para a implementação dessas tecnologias em larga escala.



Figura 2. Distribuição geográfica das publicações

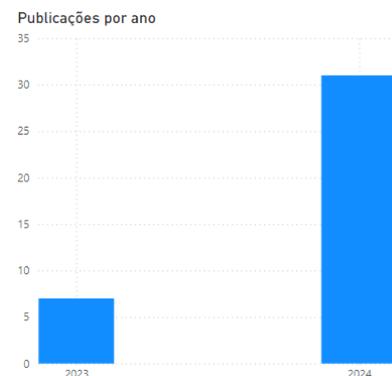


Figura 3. Publicações por Ano

### 3.1. QP1 - Quais os tipos de estudos e abordagens utilizadas?

A análise dos estudos incluídos neste mapeamento sistemático revelou uma ampla variedade de abordagens metodológicas empregadas para investigar o uso de IAs generativas na avaliação automatizada da aprendizagem. Grande parte dos trabalhos revisados utiliza estudos exploratórios e experimentais como principais metodologias. Esses estudos têm como objetivo avaliar a eficácia e a aplicabilidade de modelos de linguagem, como GPT-3 e GPT-4, em contextos educacionais, especialmente na oferta de feedback automatizado e na correção de atividades (LATIF e ZHAI, 2024; FRANKFORD *et al.*, 2024).

Estudos de caso são amplamente utilizados para examinar as experiências dos alunos com tutores virtuais baseados em IA, como ChatGPT, em diferentes disciplinas,

como programação e redação acadêmica (ABBAS; ROSMALEN e KALZ, 2023; SALEEM *et al.*, 2024). Por exemplo, em uma análise das respostas de estudantes a solicitações de ajuda em programação, verificou-se que modelos como GPT-3.5 frequentemente identificam problemas nos códigos dos alunos e oferecem explicações compreensíveis (FRANKFORD *et al.*, 2024). Essas abordagens experimentais frequentemente utilizam métricas quantitativas e qualitativas para avaliar a precisão e a utilidade das IAs generativas em comparação com os métodos tradicionais de avaliação (IBRAHIM *et al.*, 2023).

Por outro lado, algumas abordagens envolvem a aplicação de ferramentas de feedback automático em larga escala, onde o uso de IAs possibilita a escalabilidade da correção de respostas e o fornecimento de feedback qualitativo detalhado. Um exemplo é o estudo que avaliou a eficácia da IA na correção de exercícios multimodais complexos, envolvendo codificação e raciocínio econômico, mostrando que o feedback gerado pelas IAs é altamente correlacionado com o feedback de especialistas humanos (JÜRGENSMEIER e SKIERA, 2024).

Além disso, há um crescente interesse na utilização de modelos de linguagem em ambientes de aprendizado de línguas estrangeiras, como exemplificado no estudo de HÍNIZ (2024), que explora o impacto de um ano de utilização de IA generativa no ensino de línguas. Nesse cenário, tanto estudantes quanto professores relataram melhorias significativas no desenvolvimento de habilidades de escrita e comunicação.

Em suma, a diversidade de abordagens metodológicas reflete o estado ainda emergente da aplicação de IAs generativas na educação. Embora os estudos experimentais e de caso dominem o cenário, observa-se uma tendência crescente para a aplicação de metodologias quantitativas mais robustas, visando medir o impacto dessas tecnologias no desempenho e na experiência dos alunos de maneira mais sistemática e controlada (POZDNIAKOV *et al.*, 2024; POLAKOVA e IVENZ, 2024).

### 3.2. QP2 - Quais as metodologias de avaliação da aprendizagem utilizadas?

As metodologias de avaliação da aprendizagem empregadas nos estudos revisados variam amplamente, refletindo o uso diversificado de Inteligência Artificial (IA) generativa na educação. A análise dos artigos incluídos no Mapeamento Sistemático da Literatura revela que as abordagens mais comuns são:

- **Avaliação Formativa Automatizada:** Um grande número de estudos utiliza IA generativa para fornecer feedback imediato e detalhado durante o processo de aprendizagem. Essa forma de avaliação, conhecida como formativa, tem o objetivo de auxiliar o aluno a melhorar seu desempenho ao longo do curso. Exemplos incluem o uso de ferramentas como o *ChatGPT* e outros *Modelos de Linguagem de Grande Escala (LLMs)*, que avaliam textos e códigos de programação, oferecendo sugestões de melhorias ou correções automáticas. Essas abordagens são frequentemente aplicadas em disciplinas como *ciência da computação e linguagem* (MESSER *et al.*, 2024; MAHON; NAMEE e BECKER, 2024).
- **Avaliação Somativa:** Além da avaliação formativa, alguns estudos exploram a aplicação da IA em avaliações somativas, focadas na mensuração do desempenho ao final de um ciclo de aprendizagem. Esses sistemas utilizam IA para corrigir provas, testes e outras atividades de forma automatizada, proporcionando maior escalabilidade e consistência nas avaliações. Estudo de MESSER *et al.* (2024), por exemplo, analisou o uso de IA para automatizar a correção de exercícios de programação em larga escala, demonstrando como a tecnologia pode substituir ou

complementar avaliadores humanos.

- **Feedback Personalizado:** Outra metodologia importante observada nos estudos é a personalização do feedback oferecido aos alunos. Com o uso de *IAs generativas*, o feedback não apenas corrige os erros dos alunos, mas também ajusta suas respostas com base no progresso individual. FRANKFORD *et al.* (2024) mostram como o feedback pode ser ajustado em tempo real para fornecer explicações detalhadas e direcionadas, promovendo o desenvolvimento de habilidades cognitivas, como o pensamento crítico e criativo (LIANG e WU, 2024).
- **Uso de Análise de Dados Educacionais (Learning Analytics):** Muitos dos estudos revisados também utilizam a *Mineração de Dados Educacionais (EDM)* e *Learning Analytics (LA)* para avaliar o desempenho dos alunos com base em grandes volumes de dados gerados durante a interação com plataformas educacionais. Essa abordagem permite a criação de perfis de aprendizagem e o monitoramento contínuo do progresso dos estudantes, o que possibilita intervenções mais precisas. WU *et al.* (2024) discutem como esses métodos são fundamentais para fornecer uma avaliação personalizada e em tempo real, identificando rapidamente dificuldades de aprendizagem.
- **Comparação entre IA e Avaliação Humana:** Outro aspecto investigado é a comparação entre os resultados fornecidos por IAs generativas e os de avaliadores humanos. SESSLER *et al.* (2023), por exemplo, estudaram a capacidade de modelos como GPT-3.5 em identificar erros comuns em atividades experimentais e compararam esses resultados com avaliações feitas por professores, encontrando uma alta correlação em certos tipos de erro, embora as IAs ainda apresentem limitações em tarefas mais complexas.

Essas metodologias de avaliação da aprendizagem evidenciam que o uso de IAs generativas oferece soluções inovadoras para tornar a avaliação mais eficiente, escalável e personalizada. No entanto, também apontam para a necessidade de um acompanhamento rigoroso, tanto no desenvolvimento das ferramentas quanto na adaptação dos estudantes e professores às novas tecnologias.

### 3.3. QP3 - Quais as metodologias de avaliação da aprendizagem utilizadas?

Nos estudos analisados, várias técnicas de Inteligência Artificial (IA) foram utilizadas para a avaliação automatizada da aprendizagem, destacando-se as seguintes:

- **Modelos de Linguagem de Grande Escala (LLMs):** Como o *GPT-3* e *GPT-4*, que foram amplamente aplicados para a correção automática de respostas e fornecimento de feedback. Esses modelos foram explorados principalmente para disciplinas de programação e escrita (MESSER *et al.*, 2024; FRANKFORD *et al.*, 2024; MAHON; NAMEE e BECKER, 2024).
- **Análise Semântica e Similaridade de Sentenças:** Técnicas baseadas em *transformers* e *Sentence-BERT* foram utilizadas para calcular a similaridade entre respostas dos alunos e as respostas de referência, ajudando a automatizar a correção de questões abertas (RAMNARAIN-SEETOHUL; BASSOO e ROSUNALLY, 2024; SESSLER *et al.*, 2023).
- **Learning Analytics e Mineração de Dados Educacionais (EDM):** Métodos de análise de dados educacionais foram aplicados para monitorar o progresso dos alunos, identificar padrões de aprendizagem e adaptar feedback personalizado com base no desempenho contínuo (WU *et al.*, 2024; MAHON; NAMEE e BECKER, 2024).

- **Feedback Gerado por IA:** Vários estudos implementaram sistemas de feedback automatizado baseados em IA para oferecer avaliações personalizadas em tempo real, contribuindo para uma experiência de aprendizagem mais eficiente e interativa (FRANKFORD *et al.*, 2024; DAI *et al.*, 2024).

Essas técnicas são amplamente adotadas nos estudos sobre IA generativa e têm demonstrado grande potencial para melhorar a avaliação da aprendizagem, promovendo maior escalabilidade e personalização.

#### **3.4. QP4 - Foi utilizado algum outro método convencional de avaliação?**

A análise dos estudos revisados revelou que, em muitos casos, o uso de IAs generativas para a avaliação automatizada foi complementado por métodos convencionais de avaliação, como provas tradicionais, correção manual por professores e outras formas de avaliação formativa. Em particular, foi comum observar a combinação de abordagens de IA com correção manual, especialmente para validar a eficácia e a precisão das tecnologias de IA no contexto educacional (POLAKOVA e IVENZ, 2024; JÜRGENSMEIER e SKIERA, 2024).

Por exemplo, em disciplinas como programação e ciências exatas, onde a IA foi usada para corrigir automaticamente respostas de código ou problemas matemáticos, alguns estudos adotaram a revisão humana para garantir a precisão das correções e fornecer um controle de qualidade sobre os resultados fornecidos pela IA. Esse método híbrido, combinando IA e avaliação humana, demonstrou ser eficaz na identificação de erros não detectados pela IA ou na melhoria do feedback oferecido aos estudantes (FRANKFORD *et al.*, 2024).

Além disso, em estudos focados na aprendizagem de línguas, as avaliações automatizadas geradas pela IA foram frequentemente comparadas com avaliações feitas por professores. A combinação de ambos os métodos permitiu que os educadores ajustassem o feedback de forma mais precisa e garantissem que os estudantes recebessem um retorno mais completo e personalizado (HÍNIZ, 2024).

Outro método convencional frequentemente utilizado em conjunto com a avaliação automatizada foi a avaliação somativa, como exames finais e trabalhos escritos, que continuaram sendo aplicados mesmo quando a IA fornecia avaliações formativas ao longo do curso. Essa prática reflete uma tentativa de integrar o uso de IAs como uma ferramenta auxiliar, ao invés de substituírem completamente as avaliações tradicionais (POLAKOVA e IVENZ, 2024).

## **4. Conclusões**

Este estudo realizou um Mapeamento Sistemático da Literatura (MSL) sobre o uso de Inteligências Artificiais (IA) generativas na avaliação automatizada da aprendizagem, com o objetivo de responder à pergunta principal: *Como usar a Inteligência Artificial para avaliação automatizada da aprendizagem?* Com base na análise dos artigos incluídos, pode-se concluir que as IAs generativas oferecem um potencial significativo para transformar a forma como a avaliação é conduzida em ambientes educacionais, tanto na escala quanto na personalização dos feedbacks.

As IAs, especialmente modelos de linguagem de grande escala como o GPT-3 e o GPT-4, têm sido amplamente aplicadas na correção automatizada de respostas, principalmente em disciplinas como programação e redação acadêmica. Esses sistemas permitem a correção imediata e o fornecimento de feedback detalhado e adaptado às necessidades individuais dos alunos, o que potencializa o processo de ensino-aprendizagem (LATIF e ZHAI, 2024; FRANKFORD *et al.*, 2024). O uso de IAs

também tem mostrado sua eficácia na personalização do feedback, oferecendo aos alunos a oportunidade de melhorar seu desempenho de maneira contínua e em tempo real.

Além disso, o estudo revelou que as IAs podem ser usadas para complementar métodos convencionais de avaliação, como exames somativos e correção manual. Em muitos casos, a avaliação automatizada foi combinada com revisões feitas por professores, garantindo uma maior confiabilidade nos resultados e promovendo um controle de qualidade sobre as avaliações geradas pela IA (JÜRGENSMEIER e SKIERA, 2024; POLAKOVA e IVENZ, 2024). Essa abordagem híbrida tem sido particularmente útil em contextos educacionais onde a precisão é crucial, como em disciplinas de exatas e ciências da computação.

No entanto, apesar dos avanços promissores, ainda há desafios significativos que precisam ser superados para a adoção generalizada dessas tecnologias. Questões como a precisão das avaliações, a aceitação pelos educadores e estudantes, e as implicações éticas relacionadas à privacidade dos dados e à equidade de oportunidades permanecem como obstáculos importantes. A implementação em larga escala requer, além disso, uma infraestrutura tecnológica robusta e um preparo adequado dos professores para integrar essas novas ferramentas no cotidiano educacional.

Em conclusão, o uso de IA para avaliação automatizada da aprendizagem se mostra uma ferramenta poderosa para melhorar a eficiência e a qualidade da avaliação educacional. Combinando tecnologias de IA com métodos de avaliação convencionais, as instituições de ensino podem alcançar uma maior escalabilidade e personalização no acompanhamento do progresso dos alunos. No entanto, é fundamental que estudos futuros continuem a investigar maneiras de superar as barreiras associadas a essa tecnologia e explorem novas formas de torná-la mais acessível e ética no contexto educacional.

## Referências

- ABBAS, M.; ROSMALEN, P. V.; KALZ, M. A data-driven approach for the identification of features for automated feedback on academic essays. *IEEE Transactions on Learning Technologies*, IEEE, v. 16, n. 6, p. 914–928, 2023.
- CLOUDE, E. B.; KUMAR, P.; BAKER, R. S.; FOUH, E. Novice programmers inaccurately monitor the quality of their work and their peers' work in an introductory computer science course. In: *Proceedings of the 14th Learning Analytics and Knowledge Conference (LAK '24)*. [S.l.: s.n.], 2024.
- DAI, W. *et al.* Can large language models provide feedback to students? a case study on chatgpt. *Centre for Learning Analytics at Monash, Monash University*, 2024.
- FRANKFORD, E.; SAUERWEIN, C.; BASSNER, P.; KRUSCHE, S.; BREU, R. Ai-tutoring in software engineering education: Experiences with large language models in programming assessments. In: *ACM. 46th International Conference on Software Engineering: Software Engineering Education and Training (ICSE-SEET '24)*. [S.l.], 2024. p. 11.
- GÖRER, B.; AYDEMIR, F. B. Exploring the reit architecture for requirements elicitation interview training with robotic and virtual tutors. *The Journal of Systems & Software*, v. 212, p. 112018, 2024.
- HICKMAN, H.; BELL, T. Automated assessment: Does it align with teachers' views? In: *Proceedings of the 19th WiPSCE Conference on Primary and Secondary Computing Education Research*. [S.l.: s.n.], 2024.
- HÍNIZ, G. A year of generative ai in english language teaching and learning: A case study. *Journal of Research on Technology in Education*, Taylor & Francis, 2024.

- IBRAHIM, H.; LIU, F.; ASIM, R. *et al.* Perception, performance, and detectability of conversational artificial intelligence across 32 university courses. *Scientific Reports*, Nature Publishing Group, v. 13, p. 12187, 2023.
- JÜRGENSMEIER, L.; SKIERA, B. Generative ai for scalable feedback to multimodal exercises. *International Journal of Research in Marketing*, Elsevier, v. 41, p. 468–488, 2024.
- LATIF, E.; ZHAI, X. Fine-tuning chatgpt for automatic scoring. *Computers and Education: Artificial Intelligence*, Elsevier, v. 6, p. 100210, 2024.
- LIANG, W.; WU, Y. Exploring the use of chatgpt to foster efl learners' critical thinking skills from a post-humanist perspective. *Thinking Skills and Creativity*, v. 45, p. 101645, 2024.
- MAHON, J.; NAMEE, B. M.; BECKER, B. A. Guidelines for the evolving role of generative ai in introductory programming based on emerging practice. *ITiCSE 2024*, ACM, 2024.
- MESSER, M.; BROWN, N. C. C.; KÖLLING, M.; SHI, M. Automated grading and feedback tools for programming education: A systematic review. *ACM Transactions on Computing Education*, v. 24, n. 1, p. 10, 2024.
- NICKL, M. *et al.* Effects of real-time adaptivity of scaffolding: Supporting pre-service mathematics teachers' assessment skills in simulations. *Learning and Instruction*, v. 81, p. 101620, 2024.
- PAGE, M. J. *et al.* The prisma 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, v. 372, p. n71, 2021. Disponível em: (<https://doi.org/10.1136/bmj.n71>).
- PETERSEN, K.; FELDT, R.; MUJTABA, S.; MATTSSON, M. Systematic mapping studies in software engineering. In: BCS Learning & Development Ltd. *Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering (EASE'08)*. GBR, 2008. p. 68–77.
- POLAKOVA, P.; IVENZ, P. The impact of chatgpt feedback on the development of efl students' writing skills. *Cogent Education*, Taylor & Francis, v. 11, n. 1, p. 2410101, 2024.
- POZDNIAKOV, S.; BRAZIL, J.; ABDI, S. *et al.* Large language models meet user interfaces: The case of provisioning feedback. *Computers and Education: Artificial Intelligence*, Elsevier, v. 7, p. 100289, 2024.
- RAMNARAIN-SEETOHUL, V.; BASSOO, V.; ROSUNALLY, Y. Work-in-progress: Computing sentence similarity for short texts using transformer models. *University of Mauritius*, 2024.
- SALEEM, N.; MUFTI, T.; SOHAIL, S. S.; MADSEN, D. Chatgpt as an innovative heutagogical tool in medical education. *Cogent Education*, Taylor & Francis, v. 11, n. 1, p. 2332850, 2024.
- SESSLER, K.; BEWERSDORFF, A.; BAUR, A.; KASNECI, E.; NERDEL, C. Assessing student errors in experimentation using artificial intelligence and large language models: A comparative study with human raters. *Computers and Education: Artificial Intelligence*, Elsevier, 2023.
- WU, H. *et al.* Research on the construction of intelligent programming platform based on ai-generated content. *School of Computer Science and Cyber Engineering, Guangzhou University*, ACM, 2024.