

Detektoren de IA em Avaliações Acadêmicas: uma análise de confiabilidade e implicações éticas para a comunidade de IHC

Vitor Felipe B. A. Carneiro¹, Bianca C.O.E.S. Silva¹, Julia M. Rossetti¹, Luciana Correa L. F. Borges¹, Nelcileno Virgilio S. Araujo¹,

¹Instituto de Computação - Programa Pós-Graduação em Computação Aplicada –
Universidade Federal de Mato Grosso (UFMT)
Caixa Postal 78060-900 – Mato Grosso – MT – Brasil

vitorfelip@gmail.com, bianca.ces81@gmail.com, juliamariarossetti@gmail.com,
lucianafariaborges@gmail.com, nelcileno@ic.ufmt.br

Abstract. Introduction: This article discusses the risks and implications of using AI detectors in academic assessments, highlighting false positives, linguistic biases, and lack of transparency, aspects that particularly threaten early-career authors.

Objective: To critically reflect on the limitations of these detectors and propose ethical and methodological pathways for their responsible use, in dialogue with the Grand Challenges of HCI (GC2 and GC6). **Methodology:** The research was conducted through a literature review and theoretical-critical analysis, complemented by debates and internal seminars held in the Graduate Program in Applied Computing at UFMT in 2024 and 2025. **Expected Results:** This study is expected to contribute to the development of best practice guidelines, promoting greater transparency, reliability, and scientific integrity in the use of AI detectors in academic contexts.

Keywords: AI detectors, academic integrity, false positives, Human-Computer Interaction, algorithmic ethics.

Resumo. Introdução: Este artigo discute riscos e implicações do uso de detectores de IA em avaliações acadêmicas, destacando falsos positivos, vieses linguísticos e falta de transparência, aspectos que ameaçam especialmente autores em início de carreira. **Objetivo:** Refletir criticamente sobre as limitações desses detectores e propor caminhos éticos e metodológicos para seu uso responsável, em diálogo com os Grandes Desafios da IHC (GC2 e GC6). **Metodologia:** A pesquisa foi conduzida por meio de revisão bibliográfica e análise teórico-crítica, complementada por debates e seminários internos realizados no Programa de Pós-Graduação em Computação Aplicada da UFMT nos anos de 2024 e 2025. **Resultados Esperados:** Espera-se que este estudo contribua para a formulação de diretrizes de boas práticas, promovendo maior transparência, confiabilidade e integridade científica no uso de detectores de IA em contextos acadêmicos.

Palavras-Chave Detectores de IA, integridade acadêmica, falsos positivos, Intereração Humano-Computador, ética algorítmica.

1. Descrição do Desafio Abordado

Em 2024, o lançamento dos Grandes Desafios da Computação no Brasil (2025–2035) pela Sociedade Brasileira de Computação impulsionou debates sobre a necessidade de articular ética, responsabilidade e regulação no uso de tecnologias emergentes. O presente artigo insere-se nesse contexto ao tratar dos riscos e implicações do uso de detectores de Inteligência Artificial em avaliações acadêmicas, alinhando-se principalmente ao GC2: Ética e Responsabilidade e ao GC6: Implicações da Inteligência Artificial em IHC.

O GC2 enfatiza a importância da justiça procedural, da transparência e da definição de regulamentações claras para a aplicação de tecnologias, de modo a mitigar riscos associados ao seu uso. Nesse sentido, a adoção indiscriminada de detectores de IA, sem validação robusta, ameaça a integridade de processos avaliativos ao desconsiderar vieses linguísticos e a possibilidade de falsos positivos. Já o GC6 discute como a IA pode afetar práticas acadêmicas, demandando mecanismos de confiabilidade, explicabilidade e diversidade, pontos centrais para refletir sobre o impacto dos detectores em diferentes comunidades linguísticas e culturais.

O desafio dialoga ainda com questões levantadas na literatura internacional, que apontam limitações significativas na precisão desses sistemas, sobretudo em contextos de escrita acadêmica em outros idiomas. Pesquisas recentes evidenciam vieses contra autores não nativos de inglês, reforçando desigualdades já existentes no cenário científico [Liang et al 2023]. Além disso, estudos nacionais ressaltam a necessidade de discutir os limites e possibilidades dessas ferramentas, alertando para riscos éticos e metodológicos [Moraes 2024].

Dessa forma, a temática de detectores de IA em avaliações acadêmicas se apresenta como um desafio multidisciplinar, que articula ética, responsabilidade e confiabilidade tecnológica, em consonância com os Grandes Desafios da Computação no Brasil (2025–2035). O objetivo é propor reflexões críticas e caminhos para o uso responsável dessas ferramentas, assegurando maior integridade científica, respeito à diversidade e alinhamento com os princípios de Interação Humano-Computador.

2. Relatos e análises sobre o que foi realizado em 2024-2025

Entre 2024 e 2025, o Programa de Pós-Graduação em Computação Aplicada da Universidade Federal de Mato Grosso, vinculado ao Instituto de Computação, promoveu espaços de debate e reflexão acerca do uso de detectores de IA em contextos acadêmicos. Ainda que os estudos estejam em estágio inicial, seminários internos e rodas de discussão entre discentes e docentes permitiram levantar percepções sobre confiabilidade, vieses linguísticos e impactos éticos. Essas iniciativas possibilitaram uma aproximação crítica com o tema, fortalecendo a consciência coletiva sobre riscos de falsos positivos e a necessidade de protocolos de revisão humana. O processo não resultou ainda em dados científicos consolidados, mas gerou um ambiente fértil para a formulação de diretrizes preliminares e para a articulação de futuras agendas de pesquisa, alinhadas aos Grandes Desafios da IHC no Brasil.

3. Reflexões Críticas sobre as direções apontadas nos desafios

A adoção de detectores de IA por instituições acadêmicas, sem validação cuidadosa, aponta para direções que podem comprometer a integridade dos processos de avaliação. Geralmente o desenho dessas ferramentas tende a simplificar julgamentos sem apresentar classificações binárias, e sem indicar margens de erro ou níveis de incerteza, o que favorece decisões pouco transparentes e potencialmente injustas [Popkov e Barrett 2024, Moraes 2024].

Esse cenário entra em conflito com princípios consolidados da Interação Humano-Computador, que defendem confiança graduada, feedback claro e responsabilidade compartilhada [Nielsen 1994, Barbosa e Silva, 2010]. Além disso, a perda de precisão em

contextos linguísticos fora do inglês amplia vieses regionais, afetando de forma desproporcional pesquisadores brasileiros e autores em início de carreira [Liang et al 2023].

Essas limitações dialogam diretamente com os Grandes Desafios da Computação no Brasil 2025–2035, em especial os eixos sobre ética, responsabilidade e IA responsável, que enfatizam justiça procedural, transparência, explicabilidade e diversidade. Dessa forma, as reflexões críticas revelam que, sem ajustes estruturais e éticos, os detectores de IA podem reforçar desigualdades acadêmicas em vez de promover integridade científica.

4. Caminhos, estratégias e articulações para os próximos anos

Para enfrentar os riscos e limitações apontados, torna-se necessário estabelecer caminhos estratégicos que articulem boas práticas de avaliação acadêmica com os princípios da Interação Humano-Computador. Uma das primeiras medidas é a adoção de políticas de dupla revisão humana cega, garantindo que nenhuma decisão punitiva se baseie unicamente em relatórios automatizados.

Além disso, é fundamental que as ferramentas de detecção incorporem interfaces que representem incertezas de forma clara e visual, acompanhadas de justificativas compreensíveis, de modo a apoiar a interpretação crítica por parte de avaliadores. Essa abordagem fortalece a transparência e promove maior confiança nos processos de decisão.

Outro caminho estratégico está na formação continuada de docentes e avaliadores, capacitando-os para distinguir indícios preliminares de evidências conclusivas, reduzindo assim o risco de interpretações equivocadas. Também se faz necessária a criação de mecanismos transparentes de recurso, com a participação de especialistas em IHC, ética algorítmica e metodologias científicas, para assegurar decisões mais imparciais e fundamentadas.

Por fim, a articulação entre universidades, periódicos, agências de fomento e a própria comunidade de IHC é essencial para consolidar uma agenda coletiva que garanta o uso responsável dos detectores de IA, evitando práticas punitivas automáticas e fortalecendo a integridade científica.

5. Lacunas, oportunidades e parcerias

Apesar dos debates iniciais, ainda existem lacunas significativas no uso de detectores de IA em avaliações acadêmicas. Não há guias nacionais que orientem programas de pós-graduação, revistas científicas e agências de fomento quanto às condições de uso seguro e ético dessas ferramentas. Os avaliadores, em geral, não recebem treinamento específico para interpretar relatórios, o que dificulta distinguir entre indícios preliminares e provas conclusivas. Além disso, os processos de recurso permanecem pouco transparentes, raramente incluindo especialistas em IHC ou ética algorítmica.

Essas lacunas revelam oportunidades estratégicas para a comunidade de IHC. Existe espaço para liderar a elaboração de Guias de Boas Práticas, definir protocolos de uso não punitivo, propor diretrizes de comunicação transparente com os autores e recomendar capacitações periódicas para avaliadores e editores.

Nesse movimento, parcerias podem ser decisivas. A colaboração entre universidades, sociedades científicas, periódicos e órgãos de fomento pode fortalecer a legitimidade das

propostas e ampliar sua adoção. Também se abrem oportunidades de cooperação internacional, permitindo alinhar o debate a experiências globais e garantindo que os princípios de explicabilidade, confiança e justiça procedural estejam no centro da avaliação acadêmica.

6. Contribuições e reflexões para o avanço da área

Este artigo contribui para a consolidação de um debate crítico sobre os limites e riscos do uso de detectores de IA em avaliações acadêmicas. Ao problematizar falsos positivos, vieses linguísticos e a falta de transparência, o trabalho oferece uma reflexão alinhada aos princípios éticos e metodológicos da Interação Humano-Computador.

As propostas apresentadas, como a dupla revisão humana cega, a criação de interfaces mais claras e explicáveis, a formação continuada de avaliadores e os mecanismos de recurso transparente, configuram caminhos que podem orientar instituições acadêmicas rumo a práticas mais justas e confiáveis.

Além disso, ao apontar lacunas e sugerir oportunidades de cooperação, este estudo busca estimular a comunidade de IHC a assumir papel de liderança na definição de diretrizes responsáveis para o uso dessas tecnologias. Trata-se de uma contribuição que não apenas alerta para os riscos, mas também oferece bases para a construção de práticas inovadoras que fortaleçam a integridade científica, ampliem a confiança nos processos de avaliação e promovam a inclusão de diferentes contextos linguísticos e culturais.

7. Aspectos éticos envolvidos

Esta pesquisa não demandou aprovação do Comitê de Ética em Pesquisa, uma vez que foi desenvolvida exclusivamente a partir de textos científicos para revisão de literatura, em conformidade com o artigo 1º da Resolução nº 510, de 7 de abril de 2016.

8. Agradecimentos

Agradecemos o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Código de Financiamento 001. Ressaltamos que, durante a elaboração deste artigo, foram utilizados recursos tecnológicos, incluindo ferramentas baseadas em inteligência artificial, de forma exclusiva para apoio na organização e revisão textual. O emprego dessas ferramentas não substituiu a análise crítica dos autores e tampouco comprometeu o rigor científico do trabalho.

Referências

- Barbosa, S. D. J. e Silva, B. S. (2010). Livro Interação Humano-Computador. Rio de Janeiro, Elsevier.
- Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., and Zou, J. (2023). GPT detectors are biased against non-native English writers; <https://www.sciencedirect.com/science/article/pii/S2666389923001307>, Acesso em: 25 ago. 2025.

Moraes, F. S. (2024). Detecção de textos gerados pelo ChatGPT: capacidades, limitações e aplicações práticas. Revista de Ciências do Estado; <https://periodicos.ufmg.br/index.php/revice/article/view/e53509>, Acesso em: 25 ago. 2025.

Nielsen, J. (1994). Book Usability Engineering. San Francisco, Morgan Kaufmann.

Popkov, A. and Barrett, T. (2024). AI vs academia: Experimental study on AI text detectors' accuracy in behavioral health academic writing; <https://pubmed.ncbi.nlm.nih.gov/38516933/>, Acesso em: 25 ago. 2025.