

Agentes de inteligência artificial na pós-graduação *stricto sensu*: práticas, riscos e salvaguardas para integridade e autoria à luz de diretrizes ONU/UNESCO

Artificial intelligence agents in *stricto sensu* graduate programs: practices, risks, and safeguards for integrity and authorship under UN/UNESCO guidelines

Agentes de inteligencia artificial en el posgrado *stricto sensu*: prácticas, riesgos y salvaguardas para la integridad y la autoría a la luz de directrices de la ONU/UNESCO

Luís Otávio Vilela da Cruz¹

RESUMO

Este artigo analisa o uso de agentes de inteligência artificial (IA) baseados em modelos de linguagem no âmbito da pós-graduação *stricto sensu*, com ênfase em práticas acadêmicas, riscos e salvaguardas relacionadas à integridade e à autoria. Justifica-se a investigação pela intensificação do emprego desses sistemas em atividades de pesquisa e escrita, em um contexto no qual a formação do pesquisador exige rastreabilidade, responsabilidade autoral e rigor epistêmico. Metodologicamente, realizou-se pesquisa qualitativa de natureza documental e bibliográfica, com análise de conteúdo aplicada a diretrizes internacionais (ONU/UNESCO) e literatura acadêmica sobre agentes de IA, educação superior e integridade acadêmica. Os resultados indicam quatro usos recorrentes: organização do fluxo de trabalho, revisão linguística, mapeamento exploratório de literatura e apoio procedimental. Identificam-se riscos principais: fragilização da confiabilidade e da validação de fontes, reprodução de vieses e assimetrias epistêmicas, exposição indevida de dados e impactos formativos associados à delegação de operações centrais do trabalho intelectual. A discussão sistematiza diretrizes técnicas para governança do uso: transparência por etapa e finalidade, verificação obrigatória em fontes primárias, preservação de rastros mínimos, limites de não delegação para decisões analíticas e protocolos de proteção de dados. Conclui-se que a adoção responsável de agentes de IA na pós-graduação requer tradução de princípios éticos em procedimentos institucionais verificáveis, de modo a preservar autoria responsável, integridade científica e qualidade formativa.

Palavras-chave: agentes de IA; integridade acadêmica; autoria; pós-graduação *stricto sensu*; UNESCO.

ABSTRACT

¹ Mestrando no Programa de Pós-Graduação em Produção de Conteúdo Multiplataforma, Universidade Federal de São Carlos (UFSCar), ORCID: <https://orcid.org/0009-0000-8416-1319>. E-mail: luisvilelajuridico@outlook.com.

This article examines the use of large-language-model-based artificial intelligence (AI) agents in *stricto sensu* graduate programs, focusing on academic practices, risks, and safeguards related to research integrity and authorship. The study is justified by the rapid uptake of agentic systems in research and academic writing workflows within an environment that requires accountability, traceability, and epistemic rigor. Methodologically, the paper adopts a qualitative documentary and bibliographic design and applies content analysis to international guidelines (UN/UNESCO) and peer-reviewed scholarship on AI agents, higher education, and academic integrity. Findings consolidate four recurrent uses: workflow organization, language and style revision, exploratory literature mapping, and procedural support for research design. Major risks include weakened reliability and source validation, bias and epistemic asymmetries, inappropriate exposure of sensitive data, and formative impacts associated with delegating core intellectual operations. The discussion derives technical governance directions that translate ethical principles into operational rules: stage- and purpose-based transparency, mandatory verification against primary sources, minimal trace preservation, non-delegation limits for central analytic decisions, and data-protection protocols. The article concludes that responsible adoption of AI agents in graduate education depends on verifiable institutional procedures that preserve accountable authorship, research integrity, and the formative aims of postgraduate training.

Keywords: AI agents; academic integrity; authorship; *stricto sensu* graduate education; UNESCO.

RESUMEN

Este artículo analiza el uso de agentes de inteligencia artificial (IA) basados en modelos de lenguaje en el posgrado *stricto sensu*, con énfasis en prácticas académicas, riesgos y salvaguardas vinculadas a la integridad y la autoría. La investigación se justifica por la expansión de estos sistemas en tareas de investigación y escritura académica, en un ámbito que exige trazabilidad, responsabilidad autoral y rigor epistémico. Metodológicamente, se realizó un estudio cualitativo de carácter documental y bibliográfico, con análisis de contenido aplicado a directrices internacionales (ONU/UNESCO) y a literatura académica sobre agentes de IA, educación superior e integridad académica. Los resultados identifican cuatro usos recurrentes: organización del flujo de trabajo, revisión lingüística, mapeo exploratorio de literatura y apoyo procedimental al diseño de investigación. Se observan riesgos principales: debilitamiento de la confiabilidad y de la validación de fuentes, reproducción de sesgos y asimetrías epistémicas, exposición indebida de datos sensibles e impactos formativos derivados de la delegación de operaciones intelectuales centrales. La discusión sistematiza directrices técnicas para la gobernanza del uso: transparencia por etapas y finalidad, verificación obligatoria en fuentes primarias, preservación de rastros mínimos, límites de no delegación para decisiones analíticas y protocolos de protección de datos. Se concluye que la adopción responsable de agentes de IA en el posgrado requiere traducir principios éticos en procedimientos institucionales verificables, preservando la autoría responsable, la integridad científica y la calidad formativa.

Palabras clave: agentes de IA; integridad académica; autoría; posgrado *stricto sensu*; UNESCO.

1. INTRODUÇÃO

A incorporação de sistemas de inteligência artificial (IA) ao ensino superior, especialmente à pós-graduação *stricto sensu*, ganhou força com a difusão dos modelos de linguagem de grande porte. Esses modelos sustentam aplicações capazes de gerar textos, organizar informações e auxiliar tarefas acadêmicas. Quando integrados a mecanismos de planejamento, memória e uso de ferramentas, também podem compor agentes de IA que executam etapas encadeadas de um fluxo de trabalho, conforme arquiteturas de raciocínio e ação e configurações multiagentes (Yao et al., 2023; Wang et al., 2024; Guo et al., 2024). Sua presença na formação científica alcança práticas sensíveis, entre elas a autoria, a justificativa de decisões metodológicas e a rastreabilidade do percurso intelectual.

A pós-graduação *stricto sensu* articula produção de conhecimento e socialização do pesquisador para normas, valores e rotinas da comunidade científica. A formação no mestrado e no doutorado envolve a aprendizagem de critérios de qualidade, a construção de identidades acadêmicas e a participação em culturas disciplinares mediadas pelo programa e pela orientação (Weidman; Twale; Stein, 2001; Austin, 2002; Golde, 2005). Nesse ambiente, integridade acadêmica designa o compromisso verificável com honestidade, confiança, justiça, respeito, responsabilidade e coragem na produção e na comunicação do conhecimento (International Center for Academic Integrity, 2021). Sua aplicação exige fidelidade às fontes, clareza sobre contribuições, proteção dos dados, possibilidade de reconstruir decisões e disposição para responder por argumentos e resultados. A orientação participa desse processo ao acompanhar a autonomia, a originalidade e a prestação de contas do pesquisador em formação (Lee, 2008). O uso de agentes de IA em revisões de literatura, redação, organização de dados ou elaboração de instrumentos passa a integrar o regime de integridade do trabalho acadêmico.

Na literatura educacional, já se reconhecem oportunidades e desafios associados ao uso de modelos de linguagem no ensino, incluindo potencial de apoio à aprendizagem, riscos de dependência cognitiva, problemas de confiabilidade e reprodução de vieses, além de impactos sobre avaliação e autoria (Kasneci et al., 2023). No âmbito da educação em pesquisa, práticas colaborativas, como aprendizagem entre pares e produção conjunta, historicamente demandam regras explícitas de contribuição e responsabilidade (Boud; Lee, 2005). A presença de agentes de IA acrescenta um novo tipo de colaboração, não humana, cujo funcionamento depende de dados, parâmetros e instruções frequentemente opacos ao usuário final. Em consequência, torna-se necessário diferenciar usos de suporte (p. ex., revisão linguística, estruturação) de usos que podem alterar o núcleo intelectual do trabalho (p. ex., formulação de argumentos, decisões analíticas), bem como estabelecer critérios de transparência e validação.

Do ponto de vista normativo, diretrizes internacionais recentes têm reiterado princípios aplicáveis ao contexto acadêmico, como centralidade dos direitos humanos, responsabilidade, transparência, segurança, proteção de dados e prevenção de danos. A Recomendação da UNESCO sobre Ética da IA propõe

balizas para desenvolvimento e uso responsável, com atenção a impactos sociais e à necessidade de governança (UNESCO, 2021). Complementarmente, orientações específicas para IA generativa em educação e pesquisa indicam cautelas quanto à confiabilidade dos resultados, à atribuição de crédito e à necessidade de políticas institucionais que orientem práticas pedagógicas e científicas (Miao; Holmes, 2023). Em escala global, relatórios das Nações Unidas sobre governança de IA reforçam a importância de mecanismos de *accountability*, avaliação de riscos e arranjos institucionais capazes de alinhar inovação a valores públicos (United Nations, 2024). No campo da integridade acadêmica, a ênfase em honestidade, responsabilidade e justiça permanece um eixo estruturante para políticas e práticas universitárias, sobretudo em contextos de avaliação e produção científica (International Center for Academic Integrity, 2021; Bretag, 2016). Assim, há um arcabouço robusto de princípios; entretanto, persistem lacunas sobre como tais princípios se traduzem em práticas cotidianas e salvaguardas operacionais no nível dos programas de pós-graduação.

Diante desse panorama, observa-se a necessidade de pesquisas que articulem três dimensões ainda frequentemente tratadas de modo fragmentado: (i) a dimensão sociopedagógica da formação *stricto sensu* e da supervisão; (ii) a dimensão técnico-funcional dos agentes de IA (capacidade de agir por etapas, acionar ferramentas e iterar soluções); e (iii) a dimensão ético-institucional vinculada à integridade acadêmica e às diretrizes internacionais. A ausência dessa articulação dificulta delimitar usos legítimos, explicitar limites e propor procedimentos verificáveis de transparência, como registro do uso, justificativa de decisões e validação independente do conteúdo gerado.

Com base nisso, o presente artigo propõe analisar o uso de agentes de IA em atividades de pesquisa e escrita acadêmica na pós-graduação *stricto sensu*, à luz de diretrizes éticas da ONU/UNESCO e de referenciais de integridade acadêmica, buscando responder ao seguinte problema: como pós-graduandos e/ou orientadores mobilizam agentes de IA no processo de produção científica e quais salvaguardas adotam (ou deixam de adotar) para assegurar autoria responsável, confiabilidade e transparência? Ao enfrentar essa questão, espera-se contribuir para o debate acadêmico com um quadro analítico aplicável à governança do uso de IA em programas de pós-graduação e com recomendações operacionais aderentes aos princípios internacionais e às exigências de integridade científica.

2. FUNDAMENTAÇÃO TEÓRICA

A inteligência artificial generativa reúne sistemas capazes de produzir conteúdo novo, como texto, imagem ou código, a partir de padrões aprendidos em grandes conjuntos de dados. Os modelos de linguagem de grande porte, conhecidos pela sigla LLM, formam uma classe específica desses sistemas e operam pela estimativa de sequências linguísticas. Um agente de IA utiliza um modelo como componente de uma arquitetura mais ampla, na qual objetivos são decompostos, ações são selecionadas, ferramentas externas podem ser acionadas e os resultados orientam etapas posteriores (Yao et al., 2023; Wang et al., 2024).

Sistemas multiagentes reúnem dois ou mais agentes que coordenam tarefas, trocam informações ou assumem papéis distintos durante a execução. IA generativa, LLM, agente e sistema multiagente nomeiam níveis diferentes da arquitetura tecnológica, embora possam aparecer combinados em uma mesma aplicação (Guo et al., 2024). A capacidade de planejar e agir em várias etapas amplia o alcance da automação sobre o trabalho de pesquisa.

Na pós-graduação *stricto sensu*, essas arquiteturas podem apoiar leitura exploratória, síntese de literatura, planejamento de escrita, revisão linguística e organização de rotinas. Ao interferirem no encadeamento das tarefas, também influenciam escolhas, prioridades e formas de registrar o percurso intelectual. O debate educacional requer critérios que relacionem a função executada pelo sistema ao grau de responsabilidade mantido pelo pesquisador.

O estado da arte em educação superior e pós-graduação oferece um marco interpretativo relevante para compreender por que esses impactos não são periféricos. A pós-graduação tem sido analisada como um processo de socialização acadêmica no qual o estudante aprende modos de argumentar, critérios de qualidade e formas legítimas de participação na vida científica, fortemente modulados por disciplina, departamento e regimes de avaliação (Weidman; Twale; Stein, 2001; Austin, 2002; Golde, 2005). Nesse processo, a orientação assume papel estruturante, pois organiza expectativas sobre autonomia, originalidade, prestação de contas e ética na condução da pesquisa. Concepções de supervisão enfatizam que orientar não é apenas apoiar o cumprimento de etapas, mas mediar padrões de excelência, governar riscos metodológicos e sustentar um ambiente de aprendizagem em pesquisa (Lee, 2008). Quando agentes de IA passam a desempenhar funções de planejamento, sugestão de estrutura argumentativa e mesmo elaboração de versões preliminares de texto, a socialização do pesquisador pode ser afetada tanto por ganhos instrumentais (p. ex., redução de atrito na escrita) quanto por riscos formativos (p. ex., terceirização de operações cognitivas e erosão da autoria responsável), exigindo análise situada e não apenas julgamento abstrato.

A literatura recente sobre modelos de linguagem em educação, embora ainda em consolidação, já aponta oportunidades e desafios. Há evidências de potenciais usos pedagógicos relacionados à personalização de explicações, apoio à redação e ampliação de acesso a tutoria; simultaneamente, são recorrentes preocupações quanto a confiabilidade, vieses, opacidade e usos indevidos em avaliação (Kasneci *et al.*, 2023). No caso da pós-graduação, em que a escrita acadêmica é parte do próprio método de pesquisa e não um mero produto final, a distinção entre “apoio” e “substituição” torna-se menos óbvia. A questão não se reduz a detectar trechos gerados por IA, mas a compreender como o trabalho intelectual é distribuído entre humano e sistema e quais evidências permanecem disponíveis para justificar escolhas, sustentar inferências e responder por eventuais erros. Nesse ponto, o debate técnico sobre agentes converge com discussões clássicas de epistemologia prática nas Ciências Humanas, que tratam a produção de conhecimento como atividade situada, mediada por instrumentos e atravessada por relações institucionais.

Contribuições da sociologia e dos estudos sociais da ciência e tecnologia permitem enquadrar agentes de IA como componentes de redes sociotécnicas, e não como ferramentas neutras. A teoria ator-rede, por exemplo, propõe analisar a

ação como efeito de associações entre humanos e não humanos, deslocando a explicação de “intenções individuais” para arranjos que estabilizam práticas e significados (Latour, 2005). Essa chave analítica é particularmente útil para examinar o uso de agentes de IA em rotinas de pesquisa: ao sugerir caminhos, resumos e estruturas, o sistema participa da construção do problema, da organização da literatura e do modo como argumentos são encadeados. Em perspectiva próxima, Suchman (2007) discute reconfigurações humano-máquina e questiona concepções lineares de “plano” e “execução”, mostrando como a ação emerge de contingências e ajustes situados. Aplicadas ao contexto da pós-graduação, tais abordagens indicam que o uso de agentes de IA não deve ser avaliado apenas pelo resultado textual, mas pelos efeitos sobre a condução do trabalho: quais operações são automatizadas, quais decisões permanecem compreendidas pelo pesquisador e como se preserva a capacidade de justificar procedimentos diante da comunidade científica.

No campo da educação digital, há discussões que antecipam dilemas hoje reavivados pela IA generativa. Estudos sobre automação docente e “*assemblages*” pedagógicos mostram que a entrada de sistemas automatizados altera relações de ensino, regimes de presença e expectativas sobre autoridade epistêmica (Bayne, 2015). Em paralelo, análises críticas sobre dataficação e governança por dados na educação superior evidenciam que tecnologias digitais operam como dispositivos de política, reorganizando o que conta como evidência, desempenho e qualidade (Williamson, 2017). Esse repertório é valioso para compreender agentes de IA como parte de uma tendência mais ampla de reconfiguração do trabalho acadêmico por infraestruturas digitais, frequentemente orientadas por lógicas de eficiência e mensuração, mas nem sempre alinhadas aos valores formativos e democráticos da universidade.

As Ciências Humanas também oferecem um conjunto consolidado de críticas sobre poder, desigualdade e opacidade algorítmica que se torna diretamente relevante quando agentes de IA passam a mediar busca, seleção e síntese de informações. Boyd e Crawford (2012) chamam atenção para mitologias de objetividade associadas a grandes volumes de dados e para assimetrias de acesso e capacidade analítica, que criam divisões entre “ricos em dados” e “pobres em dados”. Embora agentes de IA sejam apresentados como democratização do acesso à informação, seu funcionamento depende de infraestrutura, conectividade, recursos institucionais e competências informacionais, podendo produzir novos gradientes de desigualdade entre pesquisadores e programas. Além disso, a crítica à opacidade ganha relevo: sistemas que resumem e sugerem podem obscurecer premissas, excluir dissensos e favorecer versões mais frequentes ou mais “prováveis” do discurso, afetando pluralidade teórica e diversidade epistemológica.

Nessa direção, obras de crítica sociopolítica da IA e dos algoritmos sublinham como sistemas computacionais podem reproduzir e amplificar assimetrias sociais. Noble (2018) demonstra que mecanismos de busca podem reforçar racismo e estereótipos, tornando visíveis os efeitos de hierarquias sociais na mediação algorítmica do conhecimento. O’Neil (2016) discute como modelos opacos podem produzir danos quando aplicados a decisões socialmente sensíveis, ao combinar escala, feedback e ausência de responsabilização. Eubanks (2018) evidencia como tecnologias podem intensificar vigilância e

punição sobre populações vulneráveis, oferecendo um alerta para o risco de soluções “técnicas” normalizarem desigualdades sob a aparência de neutralidade. Crawford (2021), por sua vez, desloca o debate para custos materiais e geopolíticos da IA, ao tratar seus processos como extração de recursos, trabalho e dados, com efeitos distribuídos de modo desigual no planeta. Em conjunto, essas contribuições sustentam que avaliar agentes de IA na pós-graduação exige considerar não apenas o uso local (p. ex., escrita de um capítulo), mas as condições estruturais e os efeitos de poder associados à infraestrutura tecnológica e a seus modelos de negócios e governança.

O debate crítico sobre modelos de linguagem também enfatiza riscos epistêmicos específicos: como tais sistemas geram saídas plausíveis com base em regularidades estatísticas, pode haver produção de conteúdo convincente, porém impreciso, incompleto ou sem lastro verificável, o que é particularmente problemático em formação científica. Bender *et al.* (2021) sintetizam esse problema ao alertar para perigos de modelos muito grandes, incluindo vieses, impactos ambientais e falta de transparência sobre dados de treinamento, elementos que repercutem na confiabilidade e na responsabilidade do usuário. Em Ciências Humanas, em que argumentação depende de rigor conceitual, fidelidade a fontes e explicitação de enquadramentos teóricos, o risco não está apenas em “erros factuais”, mas em distorções interpretativas, apagamento de controvérsias e homogeneização discursiva. Desse modo, o uso de agentes de IA precisa ser articulado a práticas de validação e crítica documental, preservando o princípio de que a autoria acadêmica implica responder publicamente por escolhas e inferências.

No plano normativo, diretrizes internacionais recentes oferecem balizas para orientar esse uso no ensino superior. A Recomendação da UNESCO sobre Ética da IA estabelece princípios de proteção de direitos, prevenção de danos, transparência, responsabilização e governança, indicando que aplicações de IA devem ser avaliadas por seus impactos sociais e por mecanismos institucionais de supervisão (UNESCO, 2021). Complementarmente, o guia da UNESCO para IA generativa em educação e pesquisa propõe ações e cautelas específicas para instituições, incluindo atenção a privacidade, propriedade intelectual, integridade, avaliação e capacitação docente/discente (Miao; Holmes, 2023). Em paralelo, recomendações no âmbito das Nações Unidas sobre governança da IA reforçam a necessidade de arranjos multilaterais e institucionais que promovam *accountability*, gestão de riscos e benefícios socialmente distribuídos (United Nations, 2024). Para a pós-graduação, tais diretrizes operam como uma gramática ética que pode ser traduzida em políticas locais: regras de transparência do uso, registro do apoio recebido, delimitação do que é aceitável em diferentes etapas do trabalho e mecanismos de formação para o uso crítico.

A articulação entre ética internacional e integridade acadêmica é decisiva porque títulos, publicações e avaliações dependem da confiança no processo que sustenta cada resultado. O campo da integridade acadêmica reúne valores que orientam condutas e relações institucionais: honestidade na apresentação das evidências, confiança entre os participantes, justiça na avaliação, respeito às contribuições, responsabilidade pelas decisões e coragem para enfrentar práticas inadequadas (International Center for Academic Integrity, 2021; Bretag, 2016). No trabalho com agentes de IA, esses valores se concretizam na declaração da

finalidade de uso, na conservação de versões e registros relevantes, na conferência das fontes e na assunção humana de responsabilidade pelo texto submetido. A orientação acadêmica deve incorporar esses procedimentos à formação, com critérios proporcionais à etapa da pesquisa e ao alcance da intervenção tecnológica.

Em síntese, o estado da arte indica que o uso de agentes de IA na pós-graduação deve ser compreendido como fenômeno sociotécnico e institucional: envolve arquiteturas técnicas capazes de organizar ações (Yao *et al.*, 2023; Wang *et al.*, 2024), tensões formativas no processo de socialização do pesquisador (Weidman; Twale; Stein, 2001; Austin, 2002; Golde, 2005), reconfigurações pedagógicas e de trabalho acadêmico (Bayne, 2015; Williamson, 2017), riscos de opacidade e reprodução de desigualdades (Boyd; Crawford, 2012; Noble, 2018; O’Neil, 2016; Eubanks, 2018; Crawford, 2021) e exigências normativas de governança e integridade (UNESCO, 2021; Miao; Holmes, 2023; United Nations, 2024; Bretag, 2016; International Center for Academic Integrity, 2021). O enquadramento teórico que emerge dessa convergência desloca a discussão de “ferramentas que ajudam a escrever” para “arranjos que redistribuem agência e responsabilidade”, exigindo que universidades e programas *stricto sensu* definam salvaguardas verificáveis e pedagogicamente consistentes para que a incorporação de agentes de IA não comprometa a autoria responsável nem a qualidade epistêmica da produção científica.

3. METODOLOGIA

A pesquisa caracteriza-se como qualitativa, de natureza exploratório-descritiva, desenvolvida no campo das Ciências Humanas, com delineamento de pesquisa documental e bibliográfica. A opção por esse desenho metodológico decorre da necessidade de examinar, em nível conceitual e normativo, a emergência e o uso de agentes de IA no contexto da pós-graduação *stricto sensu*, bem como as implicações éticas e de integridade acadêmica associadas a esse fenômeno, sem recorrer à coleta de dados primários com participantes humanos. A abordagem qualitativa mostra-se adequada quando o objetivo consiste em interpretar significados, marcos normativos, argumentos e disputas conceituais presentes em textos e documentos, favorecendo a reconstrução analítica do objeto no plano discursivo e institucional (Gil, 2019; Creswell; Poth, 2018).

Como materiais, foram considerados dois conjuntos principais de fontes: (a) fontes documentais normativas e institucionais, incluindo diretrizes internacionais sobre ética e governança de IA aplicáveis à educação e à pesquisa; e (b) fontes bibliográficas acadêmicas, compostas por artigos, livros e capítulos revisados por pares que tratam de (i) agentes de IA e arquiteturas baseadas em modelos de linguagem; (ii) educação superior e pós-graduação, com ênfase em socialização acadêmica, supervisão/orientação e formação para a pesquisa; e (iii) integridade acadêmica, autoria, responsabilidade e transparência no trabalho científico. Para assegurar robustez e relevância acadêmica, priorizaram-se publicações indexadas e editoras de circulação internacional reconhecida, além de documentos de organismos multilaterais diretamente vinculados ao tema, com destaque para diretrizes da UNESCO e da ONU (UNESCO, 2021; United Nations, 2024).

A seleção do corpus seguiu critérios de pertinência temática, autoridade institucional e densidade acadêmica. No eixo documental, foram incluídos documentos internacionais que explicitassem princípios, recomendações e responsabilidades relacionadas à IA, particularmente no que se refere a transparência, responsabilização, mitigação de danos, proteção de dados e implicações educacionais (UNESCO, 2021; Miao; Holmes, 2023; United Nations, 2024). No eixo bibliográfico, foram incluídos estudos que definem e caracterizam agentes de IA baseados em modelos de linguagem e suas capacidades de planejamento e ação, bem como investigações e sínteses sobre efeitos e desafios de IA generativa em contextos educacionais, com especial atenção às práticas acadêmicas de pesquisa e escrita (Yao *et al.*, 2023; Wang *et al.*, 2024; Kasneci *et al.*, 2023). Também integraram o corpus obras e estudos clássicos e contemporâneos sobre pós-graduação e formação do pesquisador, por discutirem a socialização acadêmica, a centralidade da orientação e os regimes de avaliação e autoria que estruturam a vida científica (Weidman; Twale; Stein, 2001; Austin, 2002; Golde, 2005; Lee, 2008). Por fim, foram incorporados referenciais de integridade acadêmica por sistematizarem valores e parâmetros para discussão de autoria responsável e de prestação de contas na produção científica (Bretag, 2016; International Center for Academic Integrity, 2021).

A técnica de análise adotada foi a análise de conteúdo em perspectiva qualitativa, aplicada aos documentos e textos selecionados, com vistas à identificação de unidades de sentido e à construção de inferências interpretativas sobre como o uso de agentes de IA é enquadrado em termos de riscos, benefícios, deveres e limites no contexto da pós-graduação. O procedimento analítico compreendeu três movimentos articulados: (i) leitura flutuante e delimitação do corpus, visando reconhecer a estrutura argumentativa e os principais conceitos mobilizados nas fontes; (ii) codificação e categorização, combinando categorias dedutivas, derivadas de princípios recorrentes em diretrizes internacionais (por exemplo: transparência, responsabilização, confiabilidade, privacidade, prevenção de danos), e categorias indutivas, emergentes da recorrência e do contraste entre os textos; e (iii) tratamento interpretativo dos resultados, com síntese analítica orientada pela pergunta de pesquisa e pela comparação entre os eixos documental e bibliográfico. A análise de conteúdo foi escolhida por oferecer um repertório consolidado de procedimentos para examinar comunicações e documentos, com atenção a recorrências, silêncios, tensões e deslocamentos conceituais (Bardin, 2016).

Quadro 1 – Síntese da operacionalização metodológica

Etapas	Fontes e unidades de análise	Categorias	Procedimento analítico
Constituição do corpus	Documentos ONU/UNESCO; estudos sobre agentes e LLM; pesquisas sobre pós-graduação, autoria e integridade.	Pertinência temática; autoridade institucional; densidade acadêmica; atualidade para IA.	Leitura exploratória, registro bibliográfico e exclusão de fontes sem relação direta com a pergunta de pesquisa.
Codificação dedutiva	Trechos que formulam princípios, deveres, riscos e orientações institucionais.	Transparência; responsabilização; confiabilidade; privacidade; prevenção de danos.	Marcação das unidades de sentido e comparação entre fontes normativas e acadêmicas.

Etapas	Fontes e unidades de análise	Categorias	Procedimento analítico
Codificação indutiva	Passagens sobre efeitos do uso de agentes nas práticas de pesquisa e formação.	Delegação cognitiva; homogeneização discursiva; erosão formativa; assimetrias de acesso.	Agrupamento por recorrência, contraste e relação com leitura, escrita, análise e autoria.
Síntese interpretativa	Matriz composta pelas categorias dedutivas e indutivas.	Práticas; riscos; salvaguardas; implicações para integridade e autoria.	Triangulação dos eixos do corpus e formulação de recomendações técnicas.

Fonte: elaboração própria com base no corpus da pesquisa.

O Quadro 1 sintetiza a passagem da seleção das fontes à construção das recomendações. Para fortalecer a consistência do estudo, adotou-se triangulação entre documentos normativos e de governança, literatura sobre agentes de IA e estudos sobre pós-graduação e integridade acadêmica. O confronto permitiu localizar convergências, divergências e silêncios entre os três eixos. O recorte temporal privilegiou trabalhos recentes sobre IA generativa e agentes, sem excluir obras consolidadas cuja contribuição permanece necessária à análise da formação do pesquisador (Weidman; Twale; Stein, 2001; Austin, 2002; Bretag, 2016).

Por se tratar de pesquisa exclusivamente documental e bibliográfica, sem coleta de dados com seres humanos, não se aplicam procedimentos de consentimento informado ou tramitação em comitê de ética. Ainda assim, o estudo observa princípios de integridade científica na condução da revisão e da análise, com rastreabilidade das fontes, explicitação de critérios de seleção e cuidado para não extrapolar inferências para além do que os documentos e estudos sustentam. As citações no texto seguem o sistema autor-data, e as referências são apresentadas ao final conforme norma vigente, assegurando padronização e verificabilidade.

4. RESULTADOS E DISCUSSÃO

A análise documental e bibliográfica caracterizou os agentes de IA como infraestrutura sociotécnica capaz de reorganizar o trabalho intelectual na pós-graduação *stricto sensu*. Planejamento, busca, síntese, redação e revisão podem ser incorporados a fluxos que combinam geração textual e execução orientada a objetivos (Yao et al., 2023; Wang et al., 2024). Essa mediação alcança a socialização do pesquisador e as normas de autoria e responsabilização que sustentam o campo acadêmico (Weidman; Twale; Stein, 2001; Austin, 2002; Lee, 2008).

No plano das práticas, o estado da arte indica quatro usos recorrentes e relativamente “normalizáveis” quando submetidos a salvaguardas. O primeiro é o apoio à organização do trabalho (planejamento de cronogramas, estruturação de capítulos e listas de verificação), em que o agente atua como orquestrador de tarefas, sem, necessariamente, produzir conteúdos substantivos. O segundo é o suporte à revisão linguística e coesão textual, com foco em clareza, padronização e ajuste de registro acadêmico, prática que, em princípio, pode ser aproximada de formas tradicionais de revisão, desde que não altere o núcleo argumentativo nem substitua decisões teóricas do autor. O terceiro é o uso para mapeamento

exploratório de literatura (geração de termos de busca, organização temática preliminar, sínteses iniciais), que tende a ser percebido como ganho de eficiência, mas cujo risco aumenta quando a síntese substitui a leitura efetiva e a verificação de fontes. O quarto é o apoio à modelagem procedimental (roteiros de entrevista, categorias iniciais, matrizes analíticas), que pode reduzir atritos no desenho metodológico, porém exige alto controle intelectual, por incidir diretamente sobre escolhas de pesquisa.

Essas práticas reconfiguram a relação entre pesquisador e instrumentos, o que as Ciências Humanas ajudam a explicar. Ao tratar a produção de conhecimento como ação distribuída, abordagens sociotécnicas indicam que agentes de IA se tornam participantes da rede de produção textual e decisória, com capacidade de orientar percursos e estabilizar soluções, afetando o que é considerado justificável e aceitável no processo (Latour, 2005; Suchman, 2007). Nessa chave, a questão central não é se “houve uso”, mas qual foi o grau de delegação e quais rastros foram preservados para sustentar a responsabilidade epistêmica do autor diante da comunidade científica.

No plano dos riscos, o corpus analisado converge para quatro famílias. A primeira é o risco epistêmico e de confiabilidade, associado à produção de respostas plausíveis sem lastro verificável e à tendência de homogeneização discursiva, com apagamento de controvérsias e simplificações indevidas. Esse risco é amplificado quando agentes são usados como substitutos de leitura e de crítica de fontes, ou quando se aceita como “referência” aquilo que é apenas uma geração textual (Bender et al., 2021; Kasneci et al., 2023). A segunda família é o risco de vieses e injustiças na mediação do conhecimento, pois o conteúdo e as sugestões podem reproduzir assimetrias históricas e perspectivas dominantes, afetando a diversidade epistemológica e a pluralidade teórica valorizadas nas Ciências Humanas (Boyd; Crawford, 2012; Noble, 2018; Crawford, 2021). A terceira envolve risco de privacidade, confidencialidade e proteção de dados, especialmente quando materiais sensíveis (dados de pesquisa, relatos de participantes, textos inéditos) são inseridos em sistemas com termos de uso e políticas de retenção não transparentes para o usuário (UNESCO, 2021; Miao; Holmes, 2023). A quarta família, particularmente relevante para a pós-graduação, é o risco formativo, pois a automação de operações centrais (argumentação, síntese, organização conceitual) pode reduzir oportunidades de desenvolvimento de competências acadêmicas e de autoria, afetando o processo de socialização do pesquisador e a aprendizagem das normas do campo (Weidman; Twale; Stein, 2001; Lee, 2008).

A discussão sobre riscos conecta-se diretamente à integridade acadêmica e autoria. O corpus mobilizado indica que integridade não se esgota em evitar fraude, mas envolve manter padrões de honestidade, responsabilidade e justiça no modo de produzir e reportar conhecimento, incluindo clareza sobre contribuições e rastreabilidade das decisões (Bretag, 2016; International Center for Academic Integrity, 2021). Nesse contexto, agentes de IA tensionam três dimensões: (i) atribuição de autoria, porque trechos e estruturas podem ser co-produzidos pelo sistema; (ii) responsabilidade por erros, porque o conteúdo gerado pode conter imprecisões; e (iii) transparência do processo, porque a mediação algorítmica pode ocultar como se chegou a uma formulação. Uma consequência analítica importante é que a pergunta “quem escreveu?” torna-se

menos produtiva do que “quem responde pelo argumento, pelas fontes, pelas escolhas metodológicas e pelos resultados?”. Mantém-se, assim, um princípio: ainda que haja apoio instrumental, a responsabilidade acadêmica permanece humana, pois é o pesquisador que assina, submete e defende o trabalho.

A partir das categorias dedutivas (transparência, responsabilização, confiabilidade, privacidade, prevenção de danos) e das categorias indutivas emergentes (delegação cognitiva, homogeneização discursiva, erosão do processo formativo, assimetrias de acesso), foi possível sistematizar salvaguardas com função técnica. No plano da transparência, as diretrizes internacionais recomendam que instituições e usuários adotem práticas de explicitação do uso e de governança, com especial atenção ao contexto educacional e de pesquisa (UNESCO, 2021; Miao; Holmes, 2023). Como resultado analítico, propõe-se que a transparência seja operacionalizada por (a) declaração no trabalho (ou em documento suplementar) sobre em quais etapas houve uso de agente e com qual finalidade; (b) preservação de registros mínimos (prompts relevantes, versões e decisões de revisão) para auditoria acadêmica, quando solicitado; e (c) distinção explícita entre apoio de linguagem/estrutura e contribuições substantivas. Essa operacionalização é coerente com valores de integridade, pois permite avaliar contribuição e responsabilidade sem reduzir o debate a um binarismo “usou/não usou” (International Center for Academic Integrity, 2021).

No plano da confiabilidade e validade, a análise do corpus sustenta que o uso aceitável de agentes em pós-graduação depende de uma disciplina metodológica de verificação. Recomenda-se, como diretriz técnica, que toda afirmação factual ou teórica mobilizada pelo agente seja submetida a checagem em fontes primárias (artigos, livros, documentos oficiais), com exigência de leitura e interpretação pelo pesquisador, evitando “citações por procuração”. Em Ciências Humanas, isso inclui cuidado com conceitos polissêmicos e tradições teóricas divergentes, que agentes tendem a nivelar (Bender et al., 2021). Outra salvaguarda é a regra de não delegação de decisões analíticas centrais: o agente pode sugerir categorias, mas a codificação e a justificação interpretativa devem permanecer explicitamente humanas, com base no corpus e em referenciais assumidos, sob pena de fragilizar a responsabilidade epistêmica do autor (Bardin, 2016). A recomendação é compatível com a própria ideia de supervisão como mediação de padrões e justificativas, não como terceirização de julgamento científico (Lee, 2008).

No plano de privacidade e proteção de dados, a convergência entre diretrizes da UNESCO e recomendações de governança em âmbito ONU indica a necessidade de reduzir exposição indevida de materiais, sobretudo quando há dados sensíveis, textos inéditos ou informações identificáveis (UNESCO, 2021; United Nations, 2024; Miao; Holmes, 2023). A implicação técnica para programas *stricto sensu* é a adoção de políticas de “dados mínimos”: (a) proibir ou restringir inserção de dados confidenciais em ferramentas não aprovadas institucionalmente; (b) orientar anonimização e síntese quando o apoio do agente for necessário; (c) estabelecer preferências por soluções com contratos, controles e transparência de tratamento de dados, quando disponíveis. Trata-se de uma salvaguarda de integridade científica, pois violações de confidencialidade podem comprometer pesquisa, participantes e a legitimidade do trabalho.

No plano de prevenção de danos e justiça, a literatura crítica sobre algoritmos alerta que a mediação tecnológica pode reforçar desigualdades e produzir efeitos de exclusão, inclusive na circulação do conhecimento e na formação acadêmica (Noble, 2018; O’Neil, 2016; Crawford, 2021). Em pós-graduação, isso se expressa em dois níveis: acesso diferencial a ferramentas (planos pagos, infraestrutura, proficiência) e assimetrias epistemológicas (tendência a privilegiar literatura mais visível/anglófona e perspectivas hegemônicas). Como diretriz técnica, recomenda-se que programas desenvolvam competências informacionais para uso crítico: letramento em vieses, limites de síntese, critérios de qualidade de fontes e estratégias de busca que preservem diversidade teórica e bibliográfica, evitando dependência do agente como filtro primário de relevância.

Essas salvaguardas permitem derivar um conjunto de recomendações e diretrizes para integridade e autoria no *stricto sensu*, categorizadas de forma operacional. Primeiro, no eixo autoria e contribuição, recomenda-se que o programa estabeleça um padrão de declaração de uso que diferencie: apoio linguístico/estilístico; organização estrutural; busca e triagem; apoio analítico; e geração de conteúdo. O princípio é que, quanto mais próximo do núcleo intelectual (argumento, interpretação, decisões metodológicas), maior a exigência de justificativa, rastros e supervisão. Segundo, no eixo rastreabilidade, recomenda-se que a produção acadêmica mantenha evidências do percurso (versões, principais prompts, fontes efetivamente lidas), reforçando a responsabilidade do autor e facilitando orientações formativas. Terceiro, no eixo validação, recomenda-se regra institucional de que referências e dados não podem ser “fabricados” por agentes; toda citação deve corresponder a obra real e consultada, o que endereça um risco específico de modelos de linguagem (Bender et al., 2021). Quarto, no eixo segurança e dados, recomenda-se protocolo de não inserção de dados sensíveis em sistemas não autorizados, alinhado a diretrizes internacionais de proteção e governança (UNESCO, 2021; Miao; Holmes, 2023). Quinto, no eixo formação e supervisão, recomenda-se incorporar o uso de agentes como objeto pedagógico: orientar como usá-los para ampliar capacidade crítica (comparar versões, explicitar escolhas, identificar lacunas) e não para substituir operações centrais de pesquisa, preservando a socialização acadêmica e a aprendizagem de normas do campo (Austin, 2002; Lee, 2008).

Em termos de implicações, a análise sustenta que agentes de IA tendem a acentuar uma tensão já presente na pós-graduação: a coexistência entre produtividade e qualidade epistêmica. A promessa de eficiência é real em tarefas periféricas e repetitivas; contudo, sem governança e sem literacia crítica, há risco de deslocamento do foco para resultados rápidos e de erosão do processo formativo que legitima a autoria acadêmica. Por isso, a contribuição do estudo documental-bibliográfico é indicar que a resposta institucional mais robusta não é a proibição genérica, mas a definição de limites por etapa, de critérios de transparência e de protocolos de verificação, integrados às diretrizes internacionais de ética e governança (UNESCO, 2021; United Nations, 2024). Ao traduzir princípios em regras operacionais, programas de pós-graduação podem preservar integridade e autoria responsável, ao mesmo tempo em que orientam usos legítimos de agentes de IA como suporte ao trabalho científico.

5. CONSIDERAÇÕES FINAIS

O presente artigo teve como objetivo analisar, com base em fontes documentais e bibliográficas, o uso de agentes de IA no contexto da pós-graduação *stricto sensu*, examinando práticas recorrentes, riscos associados e salvaguardas compatíveis com diretrizes éticas internacionais e com princípios de integridade acadêmica. A partir do corpus analisado, constatou-se que a adoção desses sistemas não se limita ao apoio à escrita, pois tende a reorganizar etapas do trabalho científico ao atuar na estruturação de tarefas, na mediação de informações e na padronização discursiva, com efeitos diretos sobre autoria, responsabilização e rastreabilidade intelectual.

Os achados indicam que práticas de uso mais frequentes concentram-se em organização do processo de pesquisa, revisão linguística, mapeamento exploratório de literatura e apoio procedimental. No entanto, a análise evidenciou quatro conjuntos de riscos que precisam ser tratados de modo operativo: fragilização da confiabilidade e da validação de fontes, reprodução de vieses e assimetrias epistêmicas, exposição indevida de dados e impactos formativos decorrentes de delegação de operações centrais do trabalho acadêmico. Esses riscos são particularmente relevantes nas Ciências Humanas, em que a consistência conceitual, a pluralidade teórica e a justificativa interpretativa constituem dimensões estruturantes do rigor científico.

Em resposta, o estudo sistematizou salvaguardas e diretrizes técnicas que traduzem princípios em procedimentos verificáveis. Destacam-se: transparência situada do uso por etapa e finalidade; preservação de rastros mínimos para auditoria acadêmica; verificação obrigatória em fontes primárias e rejeição de “citação por procuração”; limites de não delegação para decisões analíticas centrais; protocolos de proteção de dados e restrição ao envio de materiais sensíveis a sistemas não autorizados; e incorporação pedagógica do uso de agentes como objeto de formação crítica, articulada à supervisão. Com isso, considera-se que o trabalho atendeu ao proposto ao oferecer um quadro analítico que integra ética internacional, integridade acadêmica e características sociotécnicas dos agentes, resultando em recomendações aplicáveis à governança do uso de IA em programas de pós-graduação.

Como encaminhamentos para pesquisas futuras, recomenda-se a complementação deste estudo por investigações empíricas em programas *stricto sensu* (entrevistas, grupos focais e análise de políticas institucionais), de modo a comparar áreas, culturas disciplinares e regimes de avaliação. Sugere-se, ainda, avaliar a efetividade de diferentes formatos de declaração de uso e de protocolos de verificação, bem como explorar impactos sobre socialização acadêmica, formação de autoria e equidade de acesso a recursos tecnológicos. Essas agendas podem contribuir para consolidar políticas institucionais que preservem a integridade e a responsabilidade autoral, sem desconsiderar as transformações em curso no trabalho científico.

REFERÊNCIAS

AUSTIN, Ann E. Preparing the next generation of faculty: graduate school as socialization to the academic career. **The Journal of Higher Education**, v. 73, n.

1, p. 94-122, 2002. DOI: <https://doi.org/10.1080/00221546.2002.11777132>.

Disponível em:

<https://www.tandfonline.com/doi/abs/10.1080/00221546.2002.11777132>. Acesso em: 1 mar. 2026.

BARDIN, Laurence. **Análise de conteúdo**. São Paulo: Edições 70, 2016. ISBN 9788562938047.

BAYNE, Sian. Teacherbot: interventions in automated teaching. **Teaching in Higher Education**, v. 20, n. 4, p. 455-467, 2015. DOI:

<https://doi.org/10.1080/13562517.2015.1020783>. Disponível em:

<https://www.tandfonline.com/doi/full/10.1080/13562517.2015.1020783>. Acesso em: 3 mar. 2026.

BENDER, Emily M.; GEBRU, Timnit; McMILLAN-MAJOR, Angelina; SHMITCHELL, Shmargaret. On the dangers of stochastic parrots: can language models be too big? In: PROCEEDINGS OF THE 2021 ACM CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY (FAccT '21). **Anais [...]**. New York: ACM, 2021. p. 610-623. DOI:

<https://doi.org/10.1145/3442188.3445922>. Disponível em:

<https://dl.acm.org/doi/10.1145/3442188.3445922>. Acesso em: 5 mar. 2026.

BOUD, David; LEE, Alison. "Peer learning" as pedagogic discourse for research education. **Studies in Higher Education**, v. 30, n. 5, p. 501-516, 2005. DOI:

<https://doi.org/10.1080/03075070500249138>. Disponível em:

<https://www.tandfonline.com/doi/abs/10.1080/03075070500249138>. Acesso em: 2 mar. 2026.

BOYD, danah; CRAWFORD, Kate. Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon. **Information, Communication & Society**, v. 15, n. 5, p. 662-679, 2012. DOI:

<https://doi.org/10.1080/1369118X.2012.678878>. Disponível em:

<https://www.tandfonline.com/doi/full/10.1080/1369118X.2012.678878>. Acesso em: 4 mar. 2026.

BRETAG, Tracey (ed.). **Handbook of academic integrity**. Singapore: Springer, 2016. DOI: <https://doi.org/10.1007/978-981-287-098-8>. Disponível em:

<https://link.springer.com/referencework/10.1007/978-981-287-098-8>. Acesso em: 1 mar. 2026.

CRAWFORD, Kate. **Atlas of AI**: power, politics, and the planetary costs of artificial intelligence. New Haven: Yale University Press, 2021. Disponível em: <https://yalebooks.yale.edu/book/9780300264630/atlas-of-ai/>. Acesso em: 5 mar. 2026.

CRESWELL, John W.; POTH, Cheryl N. **Qualitative inquiry and research design**: choosing among five approaches. 4. ed. Thousand Oaks: SAGE, 2018.

Disponível em: <https://collegepublishing.sagepub.com/products/qualitative-inquiry-and-research-design-4-246896>. Acesso em: 3 mar. 2026.

EUBANKS, Virginia. **Automating inequality**: how high-tech tools profile, police, and punish the poor. New York: St. Martin's Press, 2018. ISBN 9781250074317. Disponível em: <https://us.macmillan.com/books/9781250074317/automatinginequality/>. Acesso em: 2 mar. 2026.

GIL, Antonio Carlos. **Métodos e técnicas de pesquisa social**. 7. ed. São Paulo: Atlas, 2019. ISBN 9788597020571. Disponível em: <https://www.grupogen.com.br/metodos-e-tecnicas-de-pesquisa-social/>. Acesso em: 4 mar. 2026.

GOLDE, Chris M. The role of the department and discipline in doctoral student attrition: lessons from four departments. **The Journal of Higher Education**, v. 76, n. 6, p. 669-700, 2005. DOI: <https://doi.org/10.1080/00221546.2005.11772304>. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/00221546.2005.11772304>. Acesso em: 1 mar. 2026.

GUO, Taicheng; CHEN, Xiuying; WANG, Yaqi; CHANG, Ruidi; PEI, Shichao; CHAWLA, Nitesh V.; WIEST, Olaf; ZHANG, Xiangliang. **Large language model based multi-agents**: a survey of progress and challenges. arXiv, 2024. Disponível em: <https://arxiv.org/abs/2402.01680>. Acesso em: 5 mar. 2026.

INTERNATIONAL CENTER FOR ACADEMIC INTEGRITY (ICAI). **The fundamental values of academic integrity**. 3. ed. [S. l.]: ICAI, 2021. Disponível em: <https://academicintegrity.org/aws/ICAI/pt/sp/fundamental>. Acesso em: 3 mar. 2026.

KASNECI, Enkelejda; SESSLER, Kathrin; KÜCHEMANN, Stefan; BANNERT, Maria; DEMENTIEVA, Daryna; FISCHER, Frank; GASSER, Urs; et al. ChatGPT for good? On opportunities and challenges of large language models for education. **Learning and Individual Differences**, v. 103, art. 102274, 2023. DOI: <https://doi.org/10.1016/j.lindif.2023.102274>. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S1041608023000195>. Acesso em: 2 mar. 2026.

LATOUR, Bruno. **Reassembling the social**: an introduction to actor-network-theory. Oxford: Oxford University Press, 2005. DOI: <https://doi.org/10.1093/oso/9780199256044.001.0001>. Disponível em: <https://academic.oup.com/book/52349>. Acesso em: 4 mar. 2026.

LEE, Anne. How are doctoral students supervised? Concepts of doctoral research supervision. **Studies in Higher Education**, v. 33, n. 3, p. 267-281, 2008. DOI: <https://doi.org/10.1080/03075070802049202>. Disponível em:

<https://www.tandfonline.com/doi/abs/10.1080/03075070802049202>. Acesso em: 1 mar. 2026.

MIAO, Fengchun; HOLMES, Wayne. **Guidance for generative AI in education and research**. Paris: UNESCO, 2023. DOI: <https://doi.org/10.54675/EWZM9535>. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000386693>. Acesso em: 5 mar. 2026.

NOBLE, Safiya Umoja. **Algorithms of oppression**: how search engines reinforce racism. New York: New York University Press, 2018. ISBN 9781479837243. Disponível em: <https://nyupress.org/9781479837243/algorithms-of-oppression/>. Acesso em: 3 mar. 2026.

O'NEIL, Cathy. **Weapons of math destruction**: how big data increases inequality and threatens democracy. New York: Crown, 2016. ISBN 9780553418811.

SUCHMAN, Lucy. **Human-machine reconfigurations**: plans and situated actions. 2. ed. Cambridge: Cambridge University Press, 2007. ISBN 9780521675888. Disponível em: <https://www.cambridge.org/ax/universitypress/subjects/psychology/developmental-psychology/human-machine-reconfigurations-plans-and-situated-actions-2nd-edition?format=PB&isbn=9780521675888>. Acesso em: 2 mar. 2026.

UNESCO. **Recommendation on the ethics of artificial intelligence**. Paris: UNESCO, 2021. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000380455>. Acesso em: 4 mar. 2026.

UNITED NATIONS. **Governing AI for humanity**: final report of the High-level Advisory Body on Artificial Intelligence. [S. l.]: United Nations, 2024. Disponível em: https://www.un.org/sites/un2.un.org/files/governing_ai_for_humanity_final_report_en.pdf. Acesso em: 1 mar. 2026.

WANG, Lei; MA, Chengyu; ZHANG, Xuming; et al. A survey on large language model based autonomous agents. **Frontiers of Computer Science**, v. 18, art. 186345, 2024. DOI: <https://doi.org/10.1007/s11704-024-40231-1>. Disponível em: <https://link.springer.com/article/10.1007/s11704-024-40231-1>. Acesso em: 5 mar. 2026.

WEIDMAN, John C.; TWALE, Darla J.; STEIN, Elizabeth Leahy. **Socialization of graduate and professional students in higher education**: a perilous passage? San Francisco: Jossey-Bass, 2001. (ASHE-ERIC Higher Education Report, v. 28, n. 3). Disponível em: <https://files.eric.ed.gov/fulltext/ED457710.pdf>. Acesso em: 3 mar. 2026.

WILLIAMSON, Ben. **Big data in education**: the digital future of learning, policy and practice. London: SAGE Publications Ltd, 2017. Disponível em: <https://sk.sagepub.com/book/mono/big-data-in-education/toc>. Acesso em: 2 mar. 2026.

YAO, Shunyu; ZHAO, Jeffrey; YU, Dian; DU, Nan; SHAFRAN, Izhak; NARASIMHAN, Karthik; CAO, Yuan. **ReAct**: synergizing reasoning and acting in language models. arXiv, 2023. DOI: <https://doi.org/10.48550/arXiv.2210.03629>. Disponível em: <https://arxiv.org/abs/2210.03629>. Acesso em: 4 mar. 2026.

Submetido em: 06/03/2026

Aceito em: 18/07/2026