

INTELIGÊNCIA ARTIFICIAL NAS UNIVERSIDADES: DO MÍMICO MEDIANO AO PESQUISADOR-CURADOR FUNDAMENTOS PARA UM USO RESPONSÁVEL NA PESQUISA CIENTÍFICA

ARTIFICIAL INTELLIGENCE IN UNIVERSITIES: FROM THE MEDIAN MIMIC
TO THE RESEARCHER-CURATOR FOUNDATIONS FOR RESPONSIBLE USE
IN SCIENTIFIC RESEARCH

Ciências Humanas, Engenharias • 05/07/2026

REGISTRO DOI: [10.70773/revistatopicos/782968255](https://doi.org/10.70773/revistatopicos/782968255)

Flávio Leal¹

Francis Bento Marques²

RESUMO

O artigo analisa o uso da inteligência artificial generativa nas universidades, enfatizando suas implicações epistemológicas, éticas e metodológicas para a pesquisa científica. Por meio de revisão bibliográfica exploratória e analítica, baseada em periódicos indexados na Scopus, Web of Science, SciELO e Portal de Periódicos da CAPES, o estudo organiza sua argumentação em cinco dimensões: o comportamento mimético dos grandes modelos de linguagem; as alucinações bibliográficas; os limites epistemológicos da IA quanto à originalidade e ao julgamento científico; o papel do pesquisador como curador crítico; e os princípios para o uso responsável dessas ferramentas. A tese central apoia-se no conceito de “papagaio estocástico” (Bender et al., 2021), que define os LLMs como sistemas de recombinação probabilística de padrões linguísticos, sem compreensão semântica genuína. A partir desse enquadramento, o artigo sustenta que a adoção acrítica da IA produz o “mímico mediano”: o pesquisador que delega à ferramenta a formulação de ideias e recebe conteúdos derivados do conhecimento já consolidado nos dados de treinamento. Em oposição, propõe o “pesquisador-curador”, responsável por avaliar, verificar, contextualizar e validar criticamente os resultados gerados. O estudo também apresenta uma taxonomia das principais falhas dos LLMs na produção científica, incluindo confabulação semântica, inflação retórica, presentismo bibliográfico, homogeneização estilística e supergeneralização sistemática. Conclui que a curadoria epistêmica constitui condição indispensável para preservar a integridade científica na pesquisa mediada por IA. Por fim, registra o avanço regulatório brasileiro em 2026 e propõe seis princípios: transparência, verificação sistemática, agência humana, letramento em IA, proporcionalidade e proteção de dados

Palavras-chave: Inteligência Artificial Generativa; Pesquisa científica;

Alucinação de IA; Integridade acadêmica; Curadoria do conhecimento; Ensino superior.

ABSTRACT

The article analyzes the use of generative artificial intelligence in universities, emphasizing its epistemological, ethical, and methodological implications for scientific research. Through an exploratory and analytical literature review, based on journals indexed in Scopus, Web of Science, SciELO, and the CAPES Journal Portal, the study organizes its argument into five dimensions: the mimetic behavior of large language models; bibliographic hallucinations; the epistemological limits of AI regarding originality and scientific judgment; the role of the researcher as a critical curator; and the principles for the responsible use of these tools. The central thesis is grounded in the concept of the “stochastic parrot” (Bender et al., 2021), which defines LLMs as systems of probabilistic recombination of linguistic patterns, lacking genuine semantic understanding. From this framework, the article argues that the uncritical adoption of AI produces the “mediocre mimic”: the researcher who delegates the formulation of ideas to the tool and receives content derived from knowledge already consolidated in the training data. In contrast, the study proposes the “researcher-curator,” who is responsible for critically assessing, verifying, contextualizing, and validating the generated outputs. The study also presents a taxonomy of the main failures of LLMs in scientific production, including semantic confabulation, rhetorical inflation, bibliographic presentism, stylistic homogenization, and systematic overgeneralization. It concludes that epistemic curation constitutes an indispensable condition for preserving scientific integrity in AI-mediated research. Finally, it records the advancement of the Brazilian regulatory framework in 2026 and proposes six principles:

transparency, systematic verification, human agency, AI literacy, proportionality, and data protection.

Keywords: Generative Artificial Intelligence; Scientific research; AI hallucination; Academic integrity; Knowledge curation; Higher education.

1. INTRODUÇÃO

A inteligência artificial (IAG) generativa ultrapassou as fronteiras dos laboratórios avançados e dos debates técnicos especializados com o lançamento público do ChatGPT, em novembro de 2022, instalando-se como ferramenta cotidiana nas universidades a uma velocidade que os marcos normativos, institucionais e éticos não acompanharam (Lima; Serrano, 2024). Em poucos meses, diversas ferramentas com chatbots como o Gemini (Google), o Claude (Anthropic), o Copilot (Microsoft), Perplexity AI, NotebookLM (Google) e o DeepSeek, dentre centenas de outras, passaram a frequentar salas de aula, grupos de pesquisa e bancas avaliativas de defesa, incorporadas com tal celeridade que os instrumentos regulatórios disponíveis e o letramento se revelaram estruturalmente inadequados e insuficientes para orientar, normatizar e conter os seus usos acrílicos.

O contraste entre a adoção acelerada da IA generativa e a regulação incipiente já está evidenciado em artigos da literatura especializada, que revelam a extensão e a velocidade do fenômeno. Há obviamente grandes benefícios para a pesquisa acadêmica e para os processos de ensino-aprendizagem, em todos os níveis educacionais, com o uso ético e responsável como, por exemplo: o acesso a artigos e produções científicas de centenas de países (por exemplo: IAs como Udermind, Consensus, SciSpace, Elicit, Semantic

Scholar, Research Rabbit etc.), o uso de IA com funções mais específicas como: NotebookLM que pode promover a diluição de fronteiras linguísticas, traduções, o auxílio na compreensão e estudos de textos antes inacessíveis. As IAG podem colaborar com a organização, estruturações e mapeamentos de atividades que comumente demandariam meses de trabalho como: extração de dados, organização de planilhas, correções e revisões de atividades na pesquisa que exijam protocolos já definidos previamente etc. Entretanto, embora haja esse uso em diversas áreas do conhecimento e em inúmeras profissões (Medicina, Odontologia, Engenharias etc.), no Brasil, levantamento abrangendo 158 instituições de ensino superior revelou que pouco mais de 4% delas possuíam, até 2024, qualquer posicionamento oficial sobre o uso de IA nos processos acadêmicos (Jornal da USP, 2025), demonstrando que o cenário universitário brasileiro está ainda iniciando o debate. No plano internacional, um dos primeiros levantamentos realizado pelo portal BestColleges com 1000 (mil) estudantes universitários de graduação e pós-graduação dos Estados Unidos revelou que, ao final de 2023 um ano após o lançamento do ChatGPT, 56% dos entrevistados já haviam utilizado ferramentas de inteligência artificial em atividades, pesquisa ou exames acadêmicos (Nam, 2023). Em 2026, esses números certamente são mais expressivos, pois diversos estudos recentes apontam dificuldades significativas na detecção desse uso por parte das instituições e docentes, além de abusos cometidos e percebidos na construção de pesquisas, nas atividades e nos processos de ensino-aprendizagem em todos os níveis da educação. Percebe-se, a partir desse quadro, que a questão transcende o âmbito meramente regulatório e normativo, já que, em sua dimensão mais profunda, econômica, cultural e pedagógica, trata-se de um problema epistemológico, social e educacional, que

clama por novas diretrizes, novos letramentos e novos formatos didáticos, na Educação e na Ciência.

A adoção de ferramentas de inteligência artificial (IA) por estudantes universitários atingiu, em 2025 e 2026, proporções que os principais levantamentos internacionais descrevem como próximas da universalidade. O Student Generative AI Survey 2026, publicado pelo Higher Education Policy Institute (HEPI) em parceria com a Kortext, apurou que 95% dos 1.054 graduandos britânicos entrevistados em dezembro de 2025 relatam utilizar IA de ao menos uma forma no contexto acadêmico, e 94% declaram recorrer à IA generativa para auxiliar em trabalhos submetidos a avaliação docente (Stephenson; Armstrong, 2026).

Nos Estados Unidos, o Lumina Foundation-Gallup 2026 State of Higher Education Study, conduzido com amostra de 3.801 estudantes universitários norte-americanos, igualmente ponderada por dados demográficos nacionais, corrobora esse diagnóstico de caráter irrestrito (ou de máxima abrangência no uso). O levantamento registrou que 57% dos estudantes utilizam IA em suas atividades acadêmicas ao menos semanalmente, incluindo aproximadamente um em cada cinco que declara uso diário (Marken, 2026). A frequência de uso apresenta assimetrias de gênero estatisticamente relevantes: entre os homens, 27% relatam uso diário, mas essa proporção cai para 17% entre as mulheres. As áreas de negócios (70% de uso diário ou semanal), tecnologia (68%) e engenharia (65%) concentram os usuários mais assíduos, ao passo que as humanidades apresentam os menores índices, com apenas 16% reportando uso diário e 26% declarando nunca utilizar IA nos estudos. Esses contrastes disciplinares constituem um dado estruturalmente relevante, pois sugerem que a percepção sobre a

utilidade da IA é fortemente mediada pela natureza epistemológica das práticas acadêmicas de cada campo do conhecimento e profissional.

Em território brasileiro, a pesquisa mais recente sobre o consumo e uso de inteligência artificial pela população brasileira foi realizada pelo Observatório Fundação Itaú em parceria com o Datafolha. Realizado entre os dias 07 e 15 de julho de 2025, mediante entrevistas pessoais e individuais com 2.798 pessoas a partir de 16 anos, de todas as regiões do país e classes econômicas, e margem de erro de 2 pontos percentuais para nível de confiança de 95%, o levantamento documenta que 93% dos entrevistados utilizam alguma ferramenta que aplica tecnologia de IA no cotidiano, e 75% percebem a presença da IA em seu dia a dia (OBSERVATÓRIO FUNDAÇÃO ITAÚ; DATAFOLHA, 2025). A distinção entre uso efetivo e compreensão conceitual é, contudo, marcante: 82% dos entrevistados já ouviram falar sobre IA, mas apenas 54% declaram entender o significado do termo, e 66% informam que seu conhecimento sobre o assunto provém de redes sociais. Esses dados coletados indicam que a incorporação de ferramentas de IA à rotina cotidiana dos brasileiros e do Mundo precede a formação de um referencial conceitual e de um letramento digital sobre a tecnologia adotada e usada.

Diante desse contexto socioeducacional, a tese central que orienta este artigo articula-se a partir do enquadramento conceitual proposto por Bender et al. (2021), pesquisadores que denominaram de “papagaio estocástico” (stochastic parrot) o comportamento da IA generativa, dada a sua natureza probabilística e estatística de combinações (LLMs). As IA generativas são compreendidas como um sistema que reproduz padrões linguísticos de grandes corpora

de textos sem compreensão semântica genuína sobre aquilo que enuncia e elabora, que realizam articulações aleatórias e probabilísticas de signos ao entregar resultados (outputs) medianos reprodutivos. A esse fenômeno estrutural soma-se o problema bem documentado das alucinações (hallucinations), pelo qual os modelos de linguagem fabricam referências, citações e dados inexistentes com fluência retórica convincente (Lemos, 2024; Picalho; Oliveira; Cattivelli, 2025) para responder às demandas dos seus usuários sedentos e acríticos. Além das alucinações, este estudo elenca alguns outros equívocos cometidos pelas Inteligências Artificiais: confabulações, homogeneização, inflação, presentismo etc. O pesquisador universitário que adota a IA sem a devida criticidade assume riscos nocivos capazes de comprometer a integridade do próprio trabalho científico, a sociedade e a Ciência.

O objetivo central deste artigo é oferecer ao estudante e ao pesquisador universitário, em especial ao docente-investigador das ciências humanas, sociais e da comunicação, uma análise fundamentada dos limites e das possibilidades da IA como ferramenta de pesquisa, propondo um modelo de uso responsável e ético ancorado no papel do pesquisador como curador crítico do conhecimento. A metodologia adotada é a revisão bibliográfica de caráter exploratório e analítico, com consulta a periódicos nacionais e internacionais indexados nas diversas bases Scopus, Web of Science, SciELO, REDALYC, Portal de Periódicos da CAPES etc. que se tornaram estudos analíticos, fundadores e seminais sobre essa nova realidade.

A relevância do tema no contexto do ensino superior não constitui fenômeno recente. A revisão sistemática de Zawacki-Richter et al. (2019), publicada no *International Journal of Educational Technology*

in Higher Education, analisou 146 artigos sobre aplicações de IA no ensino superior e identificou quatro domínios principais de uso: elaboração de perfis e predição; avaliação e mensuração; sistemas adaptativos e personalização; e sistemas de tutoria inteligente. Os autores alertaram, já em 2019, sobre a escassez de pesquisas que considerassem a perspectiva dos educadores no processo de adoção dessas ferramentas. O título do artigo, "onde estão os educadores?", é uma provocação central dos autores para destacar que a Inteligência Artificial na Educação (AIEd) tem sido dominada por uma perspectiva técnica em detrimento da pedagógica, lacuna que o presente estudo busca, de certa maneira, endereçar (Zawacki-Richter et al., 2019). Selwyn (2019), em obra de referência sobre IA e educação, argumenta que o entusiasmo tecnológico em torno da inteligência artificial tende a obscurecer os arranjos políticos, econômicos e pedagógicos que determinam como essas ferramentas deveriam ser efetivamente implementadas e usadas. O estudo adverte que a adoção de IA nas universidades responde com frequência a lógicas de eficiência administrativa e técnica que se chocam com os valores da educação de qualidade, da autonomia docente e do desenvolvimento do pensamento crítico (Selwyn, 2019).

Essa adoção acrítica célere dessas novas tecnologias de IA, além de gerar uma "preguiça metacognitiva" em seus usuários, modifica os processos de ensino-aprendizagem dos professores-estudantes, já que os seus estudos, trabalhos e atividades passam a ser produzidos por IA generativas sem qualquer análise crítica e leituras da agência humana, que promovem o desenvolvimento, a formação do pensamentos críticos e amadurecimentos acadêmicos inerentes ao processo de ensino-apredizagem. No âmbito da pesquisa não é diferente o cenário, pois cursos particulares sobre o uso de

ferramentas de IA surgem todos os dias em redes sociais e em instituições privadas, prometendo a construção facilitada de pesquisas, artigos, dissertações e teses em tempo recorde e com auxílio de novas tecnologias (IAs), sem qualquer responsabilidade e ética no processo da investigação científica. Diante dessa nova massificação e irresponsabilidades, cabe à Universidade promover um letramento epistêmico digital sobre o uso de Inteligências Artificiais Generativas, mas também pesquisar os impactos benéficos e prejudiciais dessas novas e excelentes ferramentas na sociedade, no ensino, na pesquisa e na extensão.

2. A IA COMO "PAPAGAIO ESTOCÁSTICO": MIMETISMO MEDIANO E CONHECIMENTO SEM COMPREENSÃO

A metáfora do “papagaio estocástico” (stochastic parrot), introduzida por Bender et al. (2021) em artigo que provocou intenso e consequente debate na comunidade acadêmica de aprendizado de máquina, oferece um dos referenciais conceituais mais precisos para compreender o funcionamento dos grandes modelos de linguagem (Large Language Models - LLMs), porque expõe a distância estrutural entre a aparência de inteligência e a ausência de compreensão, que determina a estrutura desses sistemas. A tese dos autores é direta e afirmativa. O sistema fundamentado em LLM gera estatisticamente padrões textuais a partir de volumes massivos de dados, recombina-os de forma probabilisticamente coerente, sem que isso implique compreensão, intencionalidade ou referência ao mundo. Como afirmam Bender et al. (2021, p. 616, tradução nossa):

Ao contrário do que pode parecer quando observamos seu output, um LM é um sistema para costurar aleatoriamente sequências de formas linguísticas que observou em seus vastos dados de

treinamento, de acordo com informações probabilísticas sobre como elas se combinam, mas sem qualquer referência ao significado: um papagaio estocástico.

Esse enquadramento conceitual, de alcance epistemológico considerável, revela que a aparência de inteligência, articulação, erudição e coerência dos outputs (artefatos) gerados por IA é produto de um processo estatístico e combinações, isto é, o produto não elabora compreensão, inteligência, criticidade ou raciocínio genuíno sobre o mundo. A distinção, embora técnica em sua formulação, possui implicações diretas e profundas para o uso dessas ferramentas em ambientes de pesquisa científica, pois a relação com o significado crítico e com as demandas da realidade integra as condições intrínsecas da própria atividade investigativa e científica realizada no ambiente universitário.

Santaella e Kaufman (2024), em artigo publicado na MATRIZES da Universidade de São Paulo, aprofundam essa crítica ao argumentar que os modelos de IA generativa produzem outputs linguisticamente sofisticados por meio de mecanismos estatísticos de predição, sem que lhe subjaza qualquer compreensão semântica, raciocínio causal ou experiência do mundo ordinário. Segundo as autoras, a IA generativa penetrou em território até então exclusivo da espécie humana, já que “A IA generativa, por seu lado, penetrou no segredo mais íntimo do humano: o potencial linguístico e semiótico que, até então, se constituía em marca inimitável do Sapiens” (Santaella; Kaufman, 2024, p. 46). As autoras sustentam que o advento da IA generativa representa uma “quarta ferida narcísica” para a humanidade, justamente porque o modelo, embora não compreenda o que produz, imita convincentemente atividades e habilidades humanas de linguagem e criação. Ao mesmo tempo,

ressaltam que o chatbot (interface de usuário) não opera de forma autônoma:

Sem ter de lembrar a questão óbvia de que o ChatGPT regurgita combinatórias entre bilhões de textos produzidos por humanos, o chatbot não funciona sozinho, mas implica ser acionado por usuários humanos em uma forma de diálogo linguageiro que, até então, era prerrogativa exclusiva do humano. (Santaella; Kaufman, 2024, p. 46).

Tal uso acrítico configura, portanto, um enorme risco epistêmico de natureza sistêmica (Santaella; Kaufman, 2024), cujas consequências se tornam ainda mais graves quando o pesquisador não distingue entre o que o modelo produz com aparente fluência e aquilo que, efetivamente, ele sustenta com evidências verificáveis. Quando transposta essa complexa questão para o cotidiano universitário do chão da sala de aula e do grupo de pesquisa, pensa-se em uma metáfora que ganha uma dimensão perturbadora. O usuário universitário acrítico da IA generativa comporta-se como um **mímico mediano em praça pública do conhecimento com seu papagaio repetidor**, reproduzindo com fluência e aparente erudição o denominador comum do conhecimento, já previamente registrado em sua base de treinamento. Eles, mímico e papagaio, não acessam o estado da arte, não hierarquizam evidências segundo sua qualidade epistemológica e, sobretudo, não reconhecem o que ainda não foi dito e analisado, o que é exatamente o território desconhecido em que reside a pesquisa científica original que gera impactos. No âmbito nacional, em uma obra importante e

necessária: “Diretrizes para o uso ético e responsável da Inteligência Artificial Generativa: um guia prático para pesquisadores”, Sampaio, Sabbatini e Limongi (2024, p. 13) sintetizam essa limitação estrutural ao afirmar que os grandes modelos de linguagem “apresentam semelhanças com um 'papagaio estocástico' que se esforça para replicar informações adquiridas por meio de treinamento extensivo, mas que efetivamente não compreende o que está dizendo”.

Yizhou Fan et al. (“Cuidado com a preguiça metacognitiva: efeitos da inteligência artificial generativa na motivação, nos processos e no desempenho da aprendizagem”, Beware of metacognitive laziness: effects of generative artificial intelligence on learning motivation, processes and performance, publicado no British Journal of Educational Technology, 2025) escrevem um estudo que investiga o impacto da inteligência artificial generativa, especificamente o ChatGPT, na motivação, nos processos de aprendizagem autorregulada (SRL) e no desempenho de estudantes universitários, em comparação com o apoio de especialistas humanos e com outras ferramentas tecnológicas. Os resultados expostos revelam um paradoxo significativo para analisar os ambientes dos processos de ensino-aprendizagem. Embora o grupo que utilizou a IA tenha apresentado a maior melhora imediata na pontuação de suas redações, superando inclusive o suporte de especialistas humanos experientes, esse ganho ilusório não se traduziu em um desenvolvimento real de conhecimento ou na capacidade de transferir o que foi aprendido para novos contextos. Além disso, a pesquisa demonstrou que a motivação intrínseca dos alunos permaneceu praticamente inalterada, independentemente do tipo de suporte recebido, indicando que a eficiência técnica da IA não gera necessariamente um engajamento mais profundo com o propósito da tarefa. Um dos conceitos centrais apresentados pelo

estudo é a “preguiça metacognitiva”, um fenômeno em que os alunos passam a depender excessivamente da assistência da IA, terceirizando e delegando o esforço cognitivo e o monitoramento que seriam essenciais para uma aprendizagem sólida. Enquanto a interação com tutores humanos tende a estimular processos reflexivos e conexões mais fortes entre a revisão do texto e a leitura crítica, o uso da IA pode levar ao “descarregamento cognitivo” e à “preguiça” disfarçada de eficiência, reduzindo o engajamento em estratégias de autorregulação, como a avaliação e a orientação ativa. Em última recomendação, os autores alertam que, embora a colaboração entre humanos e máquinas seja o caminho para a inteligência híbrida, é fundamental desenvolver estratégias pedagógicas, que incentivem o pensamento crítico e evitem a estagnação de habilidades humanas fundamentais diante da facilidade proporcionada pelas tecnologias generativas.

Perante essas questões em pauta e percebidas nas vivências diárias educacionais, Azambuja e Ferreira da Silva (2024), em artigo publicado na Filosofia Unisinos, argumentam que as universidades brasileiras precisam repensar suas estratégias pedagógicas à luz da IA, ressaltando que a formação universitária de qualidade exige o desenvolvimento de criatividade, pensamento crítico e competências éticas, habilidades que a fluência retórica dos modelos de linguagem é incapaz de substituir (Azambuja; Ferreira da Silva, 2024). A Universidade precisa e deve assumir a postura de formadora do pensamento crítico e reflexivo.

Trindade e Oliveira (2024), em artigo publicado nas Perspectivas em Ciência da Informação (UFMG), importante área do conhecimento para analisar esses fenômenos, investigam as habilidades informacionais necessárias ao uso de ferramentas de IA generativa

em demandas acadêmico-científicas, em estudo exploratório qualitativo. Os autores concluem que a competência em informação, sobretudo em sua dimensão ética, é condição sine qua non para o uso responsável dessas ferramentas, porque o pesquisador, sem essa competência, não consegue avaliar criticamente os outputs (produtos) gerados, tampouco verificar sua procedência ou identificar lacunas e vieses (Trindade; Oliveira, 2024). O necessário e importante letramento em IA configura-se como competência epistemológica e ética, mais do que como competência técnica no sentido restrito, conclusão essa que converge com o argumento central deste artigo.

Em editorial recente para a revista *Ensaio Pesquisa em Educação em Ciências*, Azevedo e Santos (2025) analisam a adoção acelerada da Inteligência Artificial Generativa (IAGen) na academia, identificando a emergência do que denominam “primazia da dimensão utilitária”. Este fenômeno, impulsionado por uma lógica neoliberal de produtividade instaurada, tende a privilegiar a eficiência operacional e a velocidade de produção em detrimento do rigor crítico, epistemológico e ético que constitui o cerne da produção científica legítima. Para orientar o debate, os autores sistematizam os perigos dessa adoção em cinco categorias de risco interdependentes:

- **Ilicitude:** violação de leis, normas institucionais ou direitos autorais;
- **Desconformidade:** desrespeito aos compromissos éticos e normas editoriais de transparência;

- Ilegitimidade: uso incoerente com os pressupostos científicos e apagamento de conflitos epistemológicos;
- Implausibilidade: ausência de justificativa científica sólida para a escolha do uso da tecnologia;
- Acriticidade: enfraquecimento da capacidade de problematizar estruturas de poder e comprometimento do pensamento autônomo, considerado o risco mais alarmante.

O texto sublinha que a delegação feita pelos usuários acríticos de formulação de ideias à IA conduz a uma “medianidade sistêmica”. Como os Modelos de Linguagem em Grande Escala (LLMs) operam como “papagaios estocásticos”, eles apenas repetem padrões extraídos de vastos volumes de dados, já estruturados em seus corpora. Assim, quanto mais o pesquisador renuncia de sua autoria e autonomia, mais o produto intelectual tende a um denominador comum saturado e mediano, reforçando vieses e gerando uma “monocultura de conhecimento” que mimetiza o já dito e continua reproduzindo o conhecimento com novas roupagens e embalagens. Azevedo e Santos (2025) concluem que a integridade científica não pode ser assegurada por soluções meramente tecnológicas ou normativas de controle. A resposta reside em processos formativos coletivos e no fortalecimento de uma cultura institucional de integridade, reafirmando que a ciência, ao contrário da IA, não existe para a reprodutibilidade técnica mediana, mas para o exercício da criatividade, do pensamento reflexivo e do posicionamento crítico necessários.

Esse encadeamento entre adoção irrefletida e medianidade sistêmica pode ser enunciado com alguma objetividade, pois

quanto mais o pesquisador ou o estudante delega à IA a formulação de ideias, mais o produto intelectual resultante tende ao denominador comum já saturado no corpus de treinamento do modelo. Reproduzir o já enunciado constitui, por definição, o limite estrutural da IA. Por outro lado, a pesquisa científica começa exatamente onde esse limite se impõe.

3. O MÍMICO MEDIANO: QUANDO O PAPAGAIO TEM DIPLOMA

A metáfora do “papagaio estocástico” adquire uma dimensão ainda mais alarmante, quando se percebe que ela descreve mais do que o funcionamento dos modelos de linguagem. Ela nomeia também uma patologia intelectual universitária de longa data, agravada pelas dinâmicas contemporâneas de visibilidade acadêmica e produções. O comportamento do pesquisador (ou do intelectual público) que reproduz o vocabulário, os argumentos e as posições dominantes paradigmáticas de sua comunidade de referência, sem submetê-los a exame crítico, porque essa reprodução lhe confere pertencimento, citações e reconhecimento. Esse mímico mediano opera na praça pública do conhecimento (o congresso, o grupo de discussão acadêmico, o perfil em redes de pesquisa, a mesa-redonda) da mesma forma que um LLM (Papagaio estocástico) opera sobre seu corpus de treinamento, gerando estruturas discursivas que seu público espera, recombina-as com competência retórica suficiente para produzir a impressão de erudição e maximiza a probabilidade de aprovação. O mímico mediano encontrou um papagaio que será carregado no seu ombro, sem crítica e inteligência.

A sociologia de Pierre Bourdieu oferece o instrumental analítico mais adequado para compreender conceitualmente esse fenômeno em sua dimensão institucional e de poder. No campo científico,

como em qualquer campo social, os agentes acumulam capital simbólico tanto pela contribuição original ao conhecimento quanto pela conformidade performática às regras do jogo vigentes (Bourdieu, 2004). A doxa do campo (o conjunto de pressupostos não questionados que organizam o debate legítimo) funciona como o *corpus* de treinamento do mímico, pois delimita o espaço do dizível, recompensa a reprodução elegante e penaliza a heterodoxia genuína. No contexto específico da IA generativa, esse fenômeno assume uma ironia quase didática, pois a ferramenta amplifica precisamente o mecanismo que já vigorava em algumas práticas acadêmicas menos rigorosas, existentes em muitos centros universitários e educacionais.

A consequência epistemológica é severa e nociva para todos. Uma comunidade acadêmica dominada por mímicos medianos (humanos ou artificiais) tende à estagnação disfarçada de produtividade, pois gera volume, citações cruzadas e a aparência do legítimo debate, mas não produz o desvio criativo sem o qual o conhecimento não avança e não promove desenvolvimentos. O pesquisador-curador, tal como proposto neste artigo a frente, é o antípoda do mímico mediano, que está sendo amparado e incrementado pelo papagaio. Sua conduta o distingue, já que exercita sua competência epistêmica e sua integridade ética, bem educadas, exercidas e letradas em seu compromisso político e social. Diante de qualquer nova tecnologia geradora e útil, humana ou artificial, o pesquisador-curador será sempre leitor crítico e reflexivo. A IA generativa, nesse quadro, não cria o problema do mimetismo acadêmico; ao contrário, amplifica-o e o acelera de modo sistêmico. Quando um sistema treinado para produzir um texto mais provável é adotado por agentes cujo comportamento já tende ao mais provável, o resultado é uma duplicação do mecanismo mimético. O

pesquisador-mímico usa a IA-papagaio, gerando um produto resultante que é a medianidade exponencial, fluente, referenciada, publicável, consequente e epistemologicamente oca.

4. O FENÔMENO DAS ALUCINAÇÕES: REFERÊNCIAS FANTASMAS E A FALSA ERUDIÇÃO

O fenômeno das “alucinações” (denominação técnica para o comportamento pelo qual os modelos de linguagem produzem afirmações factualmente falsas e equivocadas com aparência de fluência) constitui, entre os riscos que a IA generativa representa para a pesquisa e educação científica, o mais perigoso e traiçoeiro de todos. Esse fenômeno, decorrente de ajustes de probabilidades (estocástico), de comandos ineficientes (prompts) e de configurações do próprio sistema generativo varia bastante de intensidade, em determinadas configurações e versões dos sistemas de base generativa. Várias pesquisas analisam as manifestações desses fenômenos (Alucinações) em diversas versões de *chatbots* de Inteligências Artificiais Generativas, que variam bastante de intensidade, em áreas do conhecimento e em recorrências.

As invenções de referências, citações, conceitos e dados são extremamente nocivas à Ciência, precisamente porque sua manifestação mais insidiosa no contexto acadêmico é, por exemplo, a referência bibliográfica inventada: citações de artigos inexistentes, DOIs inválidos, autores que nunca publicaram sobre determinado tema e/ou periódicos que nunca existiram. Hicks, Humphries e Slater (2024), em estudo publicado na *Ethics and Information Technology*, discutem a natureza técnica dessa limitação por meio de um argumento filosófico rigoroso. Para esses autores, os LLMs não

mentem e “não alucinam” no sentido semântico estrito do termo, porque nem sequer pretendem representar o mundo:

O problema aqui não está em que os grandes modelos de linguagem alucinem, mintam ou deturpem o mundo de alguma maneira. O problema é que eles não foram concebidos para representar o mundo de forma alguma; foram concebidos para veicular sequências de texto convincentes. (Hicks; Humphries; Slater, 2024, p. 3, tradução nossa).

Essa caracterização expõe a raiz estrutural do problema. O LLM organiza-se pela plausibilidade perante o *corpus* de treinamento, porém a adequação à realidade permanece fora de seu alcance e de sua “inteligência”. Todos os dados irrealis são apresentados com a mesma confiança textual de uma referência legítima (Lemos, 2024; Picalho; Oliveira; Cattivelli, 2025). Esse fenômeno inerente à IA generativa é percebido mais comumente no uso de Chatbots (por exemplo: ChatGpt, Claude, Gemini, Copilot etc.), porque a interface do usuário exige um prompt (comando) para a execução de sua pesquisa e produção de produtos (output), que deve ser redigido pelo próprio usuário. Há, entretanto, IAs que delimitam seu corpus de pesquisa, evitando assim de certo modo as alucinações bibliográficas, por exemplo: Udermind, Consensus, SciSpace, Elicit, Semantic Scholar, Research Rabbit etc. porque seus *corpora* de banco de dados são limitados à pesquisa bibliográfica, evitando assim as criações de referências e dados. O Notebooklm, como uma IA “fechada” que depende de alimentação de fontes (PDFs, Vídeos,

Imagens), não comete esses equívocos, pois se limita ao que foi escolhido e alimentado (uploads) pelo usuário.

Nas IAs com chatbots, a dimensão estrutural das alucinações foi documentada com abrangência sistemática por Ji et al. (2023) em uma revisão na *ACM Computing Surveys*, que posiciona o fenômeno como um artefato intrínseco e inerente aos modelos de Geração de Linguagem (NLG), não sendo meramente uma anomalia circunstancial. Os autores sistematizam o problema em duas categorias fundamentais: alucinações intrínsecas, que apresentam contradições diretas com o conteúdo da fonte, e alucinações extrínsecas, que inserem informações não verificáveis a partir da entrada. Embora as alucinações extrínsecas possam ocasionalmente conter dados factualmente corretos do “mundo real”, elas são tratadas com cautela por representarem um risco à segurança factual e à fidelidade ao contexto fornecido. O estudo explica que essa propensão à invenção de dados é alimentada pelo viés de conhecimento paramétrico (*parametric knowledge bias*), onde os modelos priorizam padrões memorizados durante o pré-treinamento em detrimento das informações presentes na entrada de dados (input). Essa fragilidade tem implicações severas para a pesquisa científica e para domínios que exigem segurança crítica, como os médicos, onde resumos ou traduções imprecisas de instruções podem resultar em incidentes fatais ou violações de privacidade. Ji et al. enfatizam que as alucinações são particularmente insidiosas, porque o texto gerado costuma ser fluente, natural e convincente, o que “mascara” a falta de veracidade para o usuário. Esse cenário reforça a conclusão de que métricas tradicionais são inadequadas para detectar o problema, pois medem a sobreposição lexical e a fluência, mas falham em capturar a inconsistência semântica e a falta de factualidade (Ji et al.,2023).

Walters e Wilder (2023), em estudo publicado no *Scientific Reports - Nature*, ressaltam a gravidade particular desse tipo de erro para a integridade da comunicação científica:

Citações precisas fornecem evidências em sustentação às alegações, permitem aos leitores avaliar essas evidências, estabelecem o contexto para nova pesquisa empírica e chamam atenção para lacunas na literatura científica. Citações fabricadas operam contra cada um desses objetivos, sobretudo porque (a) as citações fabricadas pelo ChatGPT tendem a parecer legítimas à primeira vista e (b) o ChatGPT frequentemente apresenta respostas incorretas quando indagado 'Esta citação está correta?' ou 'Você fabrica citações?' (Walters; Wilder, 2023, p. 2, tradução nossa).

A alucinação bibliográfica ultrapassa o erro técnico corrigível. Trata-se de uma falha estrutural do modelo, que se manifesta com aparência de correção e se torna especialmente perigosa em contextos acadêmicos, onde a verificação independente deve ser hábito sistemático. O estudo de Walters e Wilder (2023) demonstra que a “alucinação bibliográfica” não é um simples deslize, mas uma consequência direta da arquitetura dos Modelos de Linguagem de Grande Escala (LLMs). Por exemplo, o ChatGPT opera como uma ferramenta de processamento de linguagem e não de processamento de informação. Ele utiliza redes neurais para prever a próxima palavra em uma sequência com base em padrões estatísticos, sendo, portanto, incapaz de distinguir entre informações

verdadeiras e falsas. Essa natureza probabilística é particularmente prejudicial para a comunicação científica por três motivos fundamentais detalhados na pesquisa: “Incompatibilidade com o Rigor Bibliográfico”, “Heurística da Confiança” (sistema é projetado para mimetizar a escrita humana e gerar feedback positivo, mesmo que “confiantemente errada”, sustentando fabricações mesmo sob questionamento); e “Mimetismo da Estrutura Científica”. O estudo identificou casos em que o ChatGPT fabricou não apenas títulos, mas também métodos, resultados e coeficientes de correlação, apresentando revisões de literatura com a estrutura de estudos empíricos inexistentes.

Os dados do estudo ilustram a persistência desse traço: enquanto o GPT-3.5 apresentou 55% de citações fabricadas, o GPT-4, embora superior, ainda manteve uma taxa de 18% de fabricações. Mesmo nas citações de obras reais, erros substantivos (como datas e números de páginas incorretos) ocorreram em 43% (GPT-3.5) e 24% (GPT-4) dos casos. Em suma, Walters e Wilder (2023) alertam que o nível de confiança depositado em softwares estatísticos é inapropriado para a IA generativa. A integridade científica exige que pesquisadores e estudantes tratem a produção da IA não como uma fonte de busca, mas como um texto que demanda verificação independente sistemática antes de qualquer uso acadêmico.

Em artigo publicado na MATRIZES da USP, Lemos (2024) analisa e classifica os comportamentos alucinatórios do ChatGPT em contextos de escrita acadêmica, demonstrando que o modelo oferece referências bibliográficas para expressões solicitadas que, na prática, não existem, erra a autoria de obras e, sobretudo, recusa-se a reconhecer que está alucinando, atribuindo seus próprios erros a

“imprecisões de informação”. Quando confrontado, admite o erro e, em seguida, reitera a alucinação ao oferecer novas fontes falsas:

Um diálogo alucinante

Você está criando uma citação falsa. O termo “alucinação da IA” não existe neste artigo.

ChatGPT – Eu peço desculpas pela confusão. Você está correto ao dizer que o termo “alucinação de IA” não é mencionado no artigo “Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks” de Radford et al. que mencionei anteriormente. Peço desculpas por qualquer confusão que eu possa ter causado. Para fornecer uma definição mais precisa de “alucinação de IA” com uma fonte, posso oferecer o seguinte: No artigo “Visualizing and Understanding Convolutional Networks” de Zeiler e Fergus, publicado na European Conference on Computer Vision (ECCV) em 2014, os autores descrevem a visualização de representações de redes neurais convolucionais (CNN) e a geração de imagens sintéticas usando o que eles se referem como um “modelo de imagem”. . . . Embora o termo “alucinação de IA” não seja usado especificamente neste artigo, esta citação destaca a ideia de uma imagem gerada por máquina que não corresponde a nenhuma cena física real, que é uma característica fundamental da alucinação de IA. (Lemos, 2024, p. 80).

O autor conclui que a alucinação da IA, no contexto específico da produção científica, pode desencadear falhas contínuas em cadeia, já que uma citação falsa incorporada sem verificação torna-se argumento, que fundamenta conclusões e alimenta outras produções subsequentes, gerando consequências difíceis de rastrear e corrigir a posteriori (Lemos, 2024).

No campo da ciência da informação, o estudo empírico de Picalho, Oliveira e Cattivelli (2025), publicado na RDBCI da Unicamp, testou comparativamente três ferramentas (ChatGPT-4o, Copilot e Gemini), na elaboração de expressões de busca em bases de dados científicos. Embora o ChatGPT tenha apresentado desempenho relativo superior, “foi possível observar erros em operadores booleanos e demais operadores avançados de pesquisa” em todas as ferramentas (Picalho; Oliveira; Cattivelli, 2025). Ainda no nível da execução técnica básica na pesquisa, portanto, a supervisão especializada se mostra indispensável.

O guia de Sampaio, Sabbatini e Limongi (2024) reforça essa posição ao precisar que os autores devem “assumir total responsabilidade pela integridade do conteúdo gerado pela IA, incluindo a revisão e a edição cuidadosa para evitar informações e citações incorretas, incompletas, inventadas ou tendenciosas” (Sampaio; Sabbatini; Limongi, 2024, p. 19). É consequência direta do princípio da autoria e da agência humana, diretriz que a proliferação de ferramentas de IA torna ainda mais urgente afirmar e exigir.

O mecanismo que produz as alucinações decorre da natureza probabilística dos LLMs como já foi ressaltado. Spinak (2023), especialista em comunicação científica da SciELO, descreve esse processo com precisão:

Os chatbots geram respostas utilizando modelos de linguagem estatísticos e baseados em probabilidades que lhes dão coerência e, em geral, são linguisticamente precisos e fluidos. No entanto, até o momento, com frequência se encontram comprometidos de várias maneiras. Por exemplo, as respostas de chatbots atualmente comportam o risco de incluir vieses, distorções, irrelevâncias, deturpações e plágio, muitos dos quais são causados pelos algoritmos que regem a sua geração e dependem, em grande medida, do conteúdo dos materiais utilizados na sua formação. (Spinak, 2023).

Ao aprofundar a explicação do mecanismo de confabulação (alucinação), Sampaio, Sabbatini e Limongi (2024) o caracterizam como resultado de uma opção estrutural dos modelos:

tais modelos podem errar ou 'mentir', inventando fatos e mesmo referências bibliográficas inexistentes! Denominado de 'confabulação' ou 'alucinação', o fenômeno ocorre por uma raridade estatística, quando a IA tenta imitar um humano, sem ter treinamento suficiente, e por terem como foco a performance, velocidade da resposta em detrimento da melhora na taxa de acerto, inventando respostas. (Sampaio; Sabbatini; Limongi, 2024, p. 14).

A partir de revisão crítica de escopo nacional e internacional, Azevedo e Santos (2025) sintetizam cinco riscos centrais identificados na literatura sobre IA generativa e integridade da pesquisa: ilicitude, desconformidade, ilegitimidade, implausibilidade e acriticidade (Azevedo; Santos, 2025). Os autores registram que os dois focos de preocupação mais recorrentes, quando a IA intervém na produção do conhecimento científico, são a inadequação na documentação de fontes e a dificuldade de verificação da veracidade das informações produzidas pelos modelos.

Peters e Chin-Yee (2025) demonstraram empiricamente esse padrão de desvio sistemático e estrutural. O estudo testou exatamente 10 modelos proeminentes: GPT-3.5 Turbo, GPT-4 Turbo, LLaMA 2 70B, Claude 2, ChatGPT-4o, ChatGPT-4.5, LLaMA 3.3 70B Versatile, Claude 3.5 Sonnet, Claude 3.7 Sonnet e DeepSeek:

A maioria dos LLMs produziu generalizações mais amplas de resultados científicos do que as contidas nos textos originais [...] os resumos gerados por LLMs foram quase cinco vezes mais propensos a conter generalizações amplas quando comparados aos resumos escritos por humanos (NEJM JW) (razão de chances [odds ratio] = 4,85, IC 95%, $p < 0,001$). (Peters; Chin-Yee, 2025, p. 12, tradução nossa).

A pesquisa conduzida por Peters e Chin-Yee (2025) consistiu em uma análise abrangente e sistemática de 4.900 resumos produzidos por dez dos principais modelos de linguagem de grande escala (LLMs), incluindo variantes do ChatGPT, Claude, LLaMA e DeepSeek.

Este estudo massivo abrangeu a sumarização de 4.300 resumos científicos e 600 artigos médicos completos provenientes de periódicos de prestígio, como Nature, Science, The Lancet e The New England Journal of Medicine. O objetivo central foi avaliar a capacidade dessas ferramentas de manter a precisão das conclusões originais, revelando que os resumos gerados por IA foram quase cinco vezes mais propensos a apresentar generalizações amplas e cientificamente indevidas quando comparados aos resumos elaborados por especialistas humanos do NEJM Journal Watch, apresentando uma razão de chances (OR) de 4,85 (IC 95%, $p < 0,001$). Essa tendência ao “viés de generalização algorítmica” manifesta-se principalmente através da omissão de nuances, limitações e quantificadores, transformando resultados específicos em afirmações genéricas ou recomendações de ação direta que os autores originais podem ter deliberadamente evitado por falta de evidências empíricas. Surpreendentemente, a pesquisa demonstrou que modelos mais recentes e potentes, como o ChatGPT-4o e o LLaMA 3.3 70B, exibiram uma propensão maior ao erro do que seus antecessores, possivelmente devido à otimização para parecerem mais prestativos e confiantes durante o treinamento com feedback humano. Em contraste, os modelos da família Claude (Anthropic) foram os únicos que não apresentaram uma diferença estatisticamente significativa, em relação aos textos originais, sugerindo que a fidelidade científica varia drasticamente conforme a arquitetura e os métodos de ajuste de cada sistema de inteligência artificial generativa.

A implicação prática é inequívoca. Toda referência e toda citação geradas por IA devem ser conferidas e verificadas de forma independente em bases de dados confiáveis (Scopus, Web of Science, PubMed, Portal de Periódicos da CAPES etc.), antes de

serem incorporadas a qualquer trabalho ou pesquisa acadêmica. Essa verificação e curadoria constituem exigência elementar de integridade científica.

5. FALHAS E EQUÍVOCOS SISTEMÁTICOS DA IAG NA PRODUÇÃO CIENTÍFICA: UMA TAXONOMIA FUNDAMENTADA

As **alucinações** bibliográficas, examinadas na seção anterior, representam apenas uma das categorias de erro que os LLMs cometem na produção científica, ainda que sejam as mais visíveis e as mais diretamente comprometedoras da integridade dos trabalhos. A literatura empírica disponível permite identificar pelo menos 08 (oito) categorias distintas de equívocos: Alucinações, Confabulações, Inflações retóricas, Falsa consensualidade, Generalizações, Presentismo, Homogeneização, perda de diversidade epistêmica coletiva etc., cada qual com mecanismo causal específico, grau de visibilidade diferenciado e implicações práticas próprias para o pesquisador universitário. Compreender essa taxonomia é condição para que o uso responsável da IA na pesquisa não produza equívocos, cujos custos epistemológicos superem os eventuais ganhos de produtividade. Reconhecer as limitações estruturais das Inteligências artificiais generativas não possui o objetivo de desqualificá-las. Por outro lado, o intuito é reconhecer a ferramenta tecnológica e suas funcionalidades para extrair o máximo de proveito, no desenvolvimento da Ciência e do próprio pesquisador.

5.1. Confabulação Semântica: Quando a Fonte Existe, Mas o Conteúdo é Distorcido

Distinta da alucinação bibliográfica, a confabulação semântica ocorre quando a referência existe, mas o conteúdo atribuído ao autor está distorcido na paráfrase e na retórica. O modelo captura o entorno textual de uma ideia sem apreender sua estrutura argumentativa real e debate, produzindo um texto que cita corretamente um artigo real, mas lhe atribui afirmações que o autor não sustenta, ou pelo menos sustenta com ressalvas que desaparecem na síntese. Farquhar et al. (2024) definem tecnicamente essa categoria:

As 'alucinações' constituem um problema crítico para os sistemas de geração de linguagem natural que utilizam grandes modelos de linguagem (LLMs) [...] Concentramo-nos em um subconjunto das alucinações que denominamos 'confabulações', no qual os LLMs formulam, com fluência, afirmações que são, ao mesmo tempo, errôneas e arbitrárias, no sentido de que a resposta é sensível a detalhes irrelevantes, como a semente aleatória (random seed). (Farquhar et al., 2024, p. 625, tradução nossa).

Os autores propõem métodos baseados em entropia semântica para detectar especificamente esse tipo de falha, já que é mais difícil de identificar justamente porque a referência existe e o texto é lido como erudição genuína (Farquhar et al., 2024). Na pesquisa filosófica e humanística, onde a precisão exegética e a fidelidade às fontes primárias são constitutivas da prática, a confabulação semântica representa um risco epistêmico de primeira grandeza, dado que o erro passa despercebido pelo leitor que não consulta o original.

5.2. Inflação Retórica e Falsa Consensualidade

A inflação retórica consiste na produção de afirmações textuais que parecem fundamentadas pela força retórica, mas carecem de base verificável e acadêmica. Expressões como “amplamente reconhecido na literatura”, “consenso científico estabelecido” e “evidências robustas indicam” são geradas pelos modelos com fluência, convencimento e frequência, independentemente de existirem ou não os estudos que as sustentariam. Uma variante desse fenômeno é a falsa consensualidade exposta pelas respostas das IAG. Ocorre quando um debate genuinamente disputado é apresentado como resolvido, consolidado e estável, porque os LLMs reproduzem o denominador comum do corpus de treinamento e tendem a omitir posições heterodoxas, minoritárias, divergentes ou em construção. Esse “viés de confirmação” já foi analisado em diversos artigos sobre as Inteligências Artificiais Generativas.

Messeri e Crockett (2024), em artigo publicado na *Nature*, formularam o conceito de “monocultura científica” para nomear esse fenômeno em escala coletiva, identificando dois subtipos e alertando para suas sérias consequências:

Uma monocultura de conhecedores emerge da priorização de um determinado ponto de vista e de todos os valores e pressupostos que ele contém, e isso influencia as perguntas de pesquisa que são formuladas e a forma como as evidências são interpretadas. As visões de IA para a ciência reforçam a ideia de um único conhecedor autoritativo e objetivo, com uma suposta “visão imparcial a partir de lugar nenhum”. As monoculturas do conhecimento surgem da priorização de uma única abordagem para formular perguntas de pesquisa e para determinar quando uma compreensão satisfatória foi alcançada, marginalizando abordagens alternativas. No contexto da pesquisa orientada por IA, uma monocultura do conhecimento pode surgir da priorização de abordagens quantitativas, reducionistas e preditivas, que são justamente aquelas que as ferramentas de IA são projetadas para otimizar. (Messeri; Crockett, 2024, p. 50, tradução nossa).

As autoras concluem com um alerta de alcance amplo: “A proliferação de ferramentas de IA na ciência corre o risco de introduzir uma fase de investigação científica na qual produzimos mais, mas compreendemos menos”. (Messeri; Crockett, 2024, p. 49, tradução nossa). O diagnóstico é de consequências graves para a diversidade epistêmica da ciência, pois uma produção acadêmica homogeneizada pela monocultura dos modelos de IA tende à estagnação mascarada de volume, gerando aparência de debate sem o avanço necessário que a heterodoxia genuína proporciona.

5.3. Presentismo Bibliográfico, Generalizações e Apagamento de Tradições Não Hegemônicas

Os LLMs tendem a priorizar a literatura recente e anglófona (Presentismo), tornando invisíveis obras fundadoras, tradições não hegemônicas (taxadas como periféricas pelo ocidente) e autores de idiomas distintos negligenciados. Há também generalizações em seus resultados que excluem divergentes e divergências. Um artigo gerado com IA pode citar dez trabalhos de 2023-2025 e ignorar a obra que originou o debate, criando aparência de atualização sem profundidade histórica. Esse viés está estruturalmente inscrito nos *corpora* de treinamento. Textos recentes em inglês são valorizados, enquanto produções em português, espanhol, mandarim, árabe e outras línguas são sistematicamente marginalizadas e desconsideradas. Essa postura ainda colonialista no século XXI perpetua nocivas posições de poder e exclusões acadêmicas, culturais, sociais e econômicas. Para os pesquisadores brasileiros e de outros países, o mecanismo implica risco adicional e concreto, pois a literatura científica nacional relevante tende a ser sub-representada ou ausente dos *outputs* gerados por IA. Cabem aos usuários das IAs generativas, ao realizarem seus prompts (comandos), solicitarem e pesquisarem as produções desses países pensantes e tão desprezados pelas academias universitárias de países imperialistas, lutando contra esse “colonialismo de poder”.

5.4. Homogeneização Estilística e Perda de Diversidade Epistêmica Coletiva

Textos gerados ou fortemente editados por IAs geralmente tendem a convergir para um mesmo padrão prosódico, de enquadramento argumentativo e de vocabulário. Doshi e Hauser (2024), em estudo

experimental publicado na *Science Advances*, demonstraram empiricamente que o acesso a LLMs melhora a criatividade individual de escritores menos criativos, mas os textos produzidos com auxílio de IA são significativamente mais semelhantes entre si do que os produzidos exclusivamente por humanos:

"[A]s histórias produzidas com o auxílio de IA generativa são mais semelhantes entre si do que aquelas escritas exclusivamente por humanos. Esses resultados apontam para um aumento da criatividade individual ao risco de perder a novidade coletiva." (Doshi; Hauser, p. 01, 2024. tradução nossa).

No plano individual, esse efeito é invisível. No plano coletivo, quando muitos pesquisadores utilizam os mesmos modelos, a produção acadêmica perde diversidade de voz e embates, de enquadramento e de perspectiva epistemológica. O Presentismo e a Homogeneização prejudicam exatamente o tipo de diversidade de que a ciência precisa para avançar além do já estabelecido e consolidado.

Embora esses resultados apontem para um aumento na criatividade individual, existe o risco de perder a novidade coletiva. [...] esse espiral descendente apresenta paralelos com um dilema social emergente: se escritores individuais descobrirem que sua escrita inspirada por IA generativa é avaliada como mais criativa, eles têm um incentivo para usar mais a IA generativa no futuro, mas, ao fazê-lo, a novidade coletiva das histórias pode ser reduzida ainda mais. (Doshi; Hauser, p. 06, 2024. tradução nossa).

5.5. A Escala Empírica do Problema: O Que os Estudos Quantitativos Mostram

As evidências empíricas sobre a magnitude dessas falhas são consistentes e bem documentadas em estudos de diferentes campos do conhecimento. Walters e Wilder (2023), em estudo publicado no Scientific Reports, produziram 84 revisões de literatura com GPT-3.5 e GPT-4, compilando 636 citações sobre 42 tópicos multidisciplinares. Os resultados indicaram que 55% das citações do GPT-3.5 eram fabricadas, assim como 18% das do GPT-4, e que 43% das referências reais do GPT-3.5 e 24% das do GPT-4 continham erros substantivos. Alguns pontos críticos foram identificados pelos pesquisadores:

1. Dificuldade com capítulos de livros: Mesmo com a evolução do modelo, ambas as versões (GPT-3.5 e GPT-4) apresentaram uma taxa de fabricação de 70% especificamente para capítulos

de livros, muitas vezes citando capítulos em livros que também são inexistentes.

2. Natureza dos erros substantivos: Nas referências que remetem a obras reais, as falhas mais frequentes ocorrem em componentes numéricos, como anos de publicação e números de volume, edição ou página. No GPT-4, embora os erros em nomes de autores e títulos sejam inferiores a 7%, a imprecisão em dados numéricos persiste como o problema principal.
3. Links enganosos: O estudo aponta que as IAs fornecem hiperlinks para menos de 10% das obras reais citadas e, quando o fazem, cerca de um terço desses links são imprecisos. Curiosamente, links são encontrados com mais frequência em citações fabricadas do que em reais.
4. O fenômeno “Confidently Wrong”: Um dos maiores perigos apontados é que o ChatGPT é “confiante no erro”, sustentando afirmações incorretas mesmo quando questionado sobre a veracidade da citação. Isso ocorre porque a ferramenta é projetada para fornecer respostas que satisfaçam o usuário, priorizando a confiança da linguagem em detrimento da precisão factual.
5. Causa fundamental: Os autores explicam que o ChatGPT não é uma ferramenta de processamento de informações, mas sim de processamento de linguagem. Ele funciona prevendo a próxima palavra em uma sequência com base em padrões estatísticos, o que o impede de distinguir entre informações precisas e falsas.

O estudo conclui que, para manter a integridade científica, pesquisadores não podem depositar nas ferramentas de IA o mesmo nível de confiança dedicado a softwares estatísticos, sendo assim indispensável a verificação manual de cada referência gerada. Como um exemplo mais notório e nocivo à Ciência, na Revista da área das Ciências Médicas, *Cureus Journal of Medical Science*, periódico médico revisado por pares e de acesso aberto (Open Access) publicado pela *Springer Nature*, em São Francisco, EUA, Bhattacharyya et al. (2023), publicando o estudo: “Altas Taxas de Referências Fabricadas e Imprecisas em Conteúdo Médico Gerado pelo ChatGPT”, encontraram taxas de “alucinações” e equívocos ainda mais expressivas em contexto das Pesquisas Médicas:

Os achados deste estudo enfatizam a necessidade de cautela na busca por informações médicas no ChatGPT, dado que a maioria das referências fornecidas mostrou-se fabricada ou imprecisa. Recomenda-se aos indivíduos que verifiquem informações médicas em fontes confiáveis e evitem confiar exclusivamente em conteúdo gerado por inteligência artificial. (Bhattacharyya et al., p. 06, 2023, tradução nossa).

O estudo de Mehul Bhattacharyya et al. (2023) detalha que a precisão das referências geradas pelo ChatGPT-3.5 variou significativamente conforme o tópico médico solicitado, revelando que em 30 artigos gerados, a taxa de referências fabricadas atingiu níveis alarmantes em certas áreas. Por exemplo, em artigos sobre disparidades na saúde, a taxa de referências totalmente inexistentes

subiu para 66%, enquanto no campo da pneumologia esse índice atingiu 75%, seguido de perto pela dermatologia com 64% e gastroenterologia com 62%. Por outro lado, a área de imunologia apresentou a menor taxa de fabricação (10%), embora a frequência de referências totalmente precisas tenha permanecido baixa em todas as categorias analisadas, representando apenas 7% do total de 115 referências avaliadas.

Para entender a gravidade das imprecisões, os autores avaliaram sete componentes específicos de cada citação e constataram que a média de componentes incorretos foi de 4,3 por referência. A média de 4,3 componentes incorretos por referência indica que, em cada citação médica produzida pelo ChatGPT, mais da metade das informações fundamentais de identificação continha erros. A gravidade dessa média decorre do fato de que a inteligência artificial raramente produz uma citação impecável, com apenas 7% das referências sendo totalmente precisas. A alta frequência de erros em campos específicos, como o PMID (93% de incorreção), volume (64%) e páginas (64%), contribuiu diretamente para elevar esse número médio de falhas por item. Além disso, como quase metade das citações (47%) foi totalmente fabricada pela IA, essas referências inexistentes acabam por acumular erros em praticamente todos os sete componentes avaliados. Em termos comparativos, essa média reflete uma imprecisão muito mais severa do que a observada em artigos científicos revisados por pares, onde as taxas de erro em citações costumam flutuar entre 4% e 48%. Os autores atribuem essa falha ao funcionamento dos modelos de linguagem de grande escala (LLMs), que utilizam redes neurais profundas para prever a próxima palavra em uma sequência com base em padrões estatísticos, não possuindo a capacidade intrínseca de distinguir entre fatos e inverdades. Isso resulta em “alucinações”, onde o sistema gera conteúdos que parecem plausíveis e são entregues de

forma convincente, mas que são, na verdade, inexistentes ou distorcidos, representando riscos reais de diagnóstico errôneo ou recomendações de tratamentos nas Ciências Médicas inadequados para o público geral.

O conjunto dessas evidências impõe uma consequência prática irrecusável e inegociável. O pesquisador que utiliza IAG na produção acadêmica sem verificação sistemática de cada *output* gerado assume riscos que a literatura empírica documenta com consistência e escala crescentes. A verificação responsável, independente e autônoma em bases de dados primárias impõe-se como exigência elementar para a integridade científica, como reforçam Hayashi (2025) e Sampaio, Sabbatini e Limongi (2024).

6. CRIATIVIDADE, ORIGINALIDADE E OS LIMITES EPISTEMOLÓGICOS DA INTELIGÊNCIA ARTIFICIAL GENERATIVA

A dimensão heurística, inovadora e original do fazer científico deve estar presente no cotidiano da Universidade, pois a ciência avança por leituras criativas. Essas características da investigação acadêmica promovem as hipóteses inesperadas, o labor intelectual epistêmico, os *insights*, as intuições, as conexões inéditas entre campos aparentemente díspares, os questionamentos de paradigmas estabelecidos que impulsionam o avanço do conhecimento. Essa inovação inerente à agência humana está estruturalmente fora da competência da IA generativa, porque, tendo em vista a própria lógica dos modelos de linguagem, essa limitação técnica estrutural não será totalmente corrigível por versões futuras. Os LLMs funcionam e trabalham sobre o que já foi dito, mapeado, criado e elaborado. Eles compilam, organizam, recombinaem e estruturam em seus *corpora* o já existente, logo seu

espaço de possibilidades está inteiramente contido no corpus histórico sobre o qual já foram treinados, ou seja, são ainda incapazes de produzir o que ainda não existe, no horizonte do já enunciado e criado.

Santaella e Kaufman (2024) articulam o limite dos modelos de linguagem com precisão filosófica ao descreverem o advento da inteligência artificial generativa como a “quarta ferida narcísica da humanidade”, evidenciando que essas ferramentas são incapazes de agência epistêmica genuína. Por serem sistemas “prisioneiros da linguagem” e destituídos do “senso comum da vida vivida”, os modelos carecem de intenções próprias, pois não formulam problemas e não possuem a autonomia necessária para escolher o que investigar, operando estritamente através da estimativa estatística da “continuação razoável” de um texto baseada em probabilidades de bilhões de páginas da web. Atribuir inteligência criativa, labor intelectual ou agenciamento epistêmico a esses sistemas é um erro, porque, conforme advertem as autoras, a IA generativa apenas mimetiza habilidades humanas por meio de simulações convincentes, mas também pode falhar em aspectos fundamentais da cognição, como abstração, raciocínio e factualidade. Essa limitação técnica é agravada pelo risco de “alucinações” e pela incapacidade do modelo em avaliar a relevância científica ou a veracidade de suas produções. Esse cenário estabelece um paradoxo de desempenho, no qual o apoio da IA mascara a erosão do pensamento crítico e da criatividade original, tornando o pesquisador dependente de uma “muleta tecnológica” que, se removida, revela uma queda acentuada na agência e na capacidade intelectual humana (Santaella; Kaufman, 2024).

Na mesma direção e argumentações sobre o reconhecimento da importância da agência e da criatividade humana epistêmica, Azambuja e Ferreira da Silva (2024) defendem que a formação universitária de qualidade deve priorizar o desenvolvimento de criatividade, pensamento crítico e competências éticas, uma vez que a fluência retórica dos modelos de linguagem, embora impressionante, é insuficiente para substituir a subjetividade humana e o julgamento moral. Esta perspectiva ganha força ao considerar que a inteligência artificial generativa (IAG) é altamente eficaz no processamento de conhecimentos explícitos e na automação de tarefas técnicas, mas falha em replicar a formação de virtudes intelectuais que têm finalidade em si mesmas, característica essencial da chamada “educação liberal”. A integração da IA generativa no aprendizado humano traz consigo o risco de uma “ilusão de desempenho”, na qual o estudante atinge resultados acadêmicos satisfatórios sem o correspondente desenvolvimento de competências cognitivas reais ou agência, o que reforça a necessidade de uma alfabetização tecnológica ou “technotização” permanente para educadores e discentes. Nesse cenário de hibridização, o papel do docente transita de um transmissor de saberes para um mentor e facilitador, responsável por desafiar a complacência e guiar os estudantes dentro de sua zona de desenvolvimento, utilizando a tecnologia como um facilitador cognitivo em vez de uma muleta que compromete a autonomia intelectual. Portanto, o desafio contemporâneo das universidades reside na elaboração de uma concepção pedagógica voltada para a formação integral e humana, inspirada no ideal clássico de kalokagathia, onde o diferencial humano reside na capacidade de fornecer a direção ética e moral para uma civilização cada vez mais mediada por sistemas que, apesar de sua eficiência técnica, operam sem autoconsciência. Segundo os autores:

A nosso ver, trata-se, de agora em diante, em certo sentido, de retorno criativo e atualizado à formação clássica proposta pelo ideal grego do desenvolvimento intelectual, crítico e criativo, com foco na educação moral e estética da juventude, a famosa kalokagathia (Azambuja e Ferreira da Silva, 2024, p. 14).

A literatura especializada documenta que a IA generativa pode apoiar dimensões operacionais do aprendizado, como a recuperação e a síntese de informações, mas evidencia limitações estruturais justamente nas competências cognitivas de ordem superior, tais como: análise crítica, síntese criativa, avaliação de evidências e formulação de problemas de pesquisa (Messerli; Crockett, 2024; Fan et al., 2025). São precisamente essas as competências que distinguem o pesquisador do mero reprodutor e compilador.

Em relação ao revés da criatividade acadêmica autoral, a desonestidade acadêmica associada ao uso de IA é bastante preocupante. Os debates contemporâneos sobre desonestidade acadêmica associada à Inteligência Artificial (IA) indicam que o problema vai além da simples autoria, atingindo a própria essência do desenvolvimento intelectual. Círico (2024) alerta que a linha entre o conhecimento autoral humano e o conteúdo produzido pela máquina torna-se perigosamente imperceptível, quando estudantes recorrem à IA para burlar o esforço exigido em atividades avaliativas e de aprendizagem, comprometendo a ética e a integridade da pesquisa, da formação e dos estudos. Este cenário é agravado pelo chamado paradoxo do desempenho, um fenômeno em que o

suporte da IA permite resultados superiores, porém cuja remoção revela uma queda brusca na capacidade real cognitiva do discente, evidenciando uma ilusão de aprendizagem que mascara a falta de desenvolvimento de habilidades como a autorregulação.

A ausência de diretrizes claras e formações éticas em instituições e periódicos científicos cria um ambiente onde os limites entre a assistência legítima e a substituição indevida da cognição humana são frequentemente ignorados e burlados. Como os modelos de linguagem conseguem mimetizar com precisão o estilo de escrita e o raciocínio humano, as avaliações tradicionais focadas apenas no produto final tornam-se ineficazes para distinguir a capacidade intelectual genuína da simulação elaborada por algoritmos. Por fim, segundo Círico (2024), a falta de alfabetização em IA entre discentes e docentes transforma uma ferramenta com alto potencial de personalização em uma “muleta” tecnológica que, em vez de expandir o conhecimento, pode comprometer a autonomia e a agência dos estudantes e pesquisadores. A superação desses desafios exige não apenas a implementação de comitês de ética e políticas editoriais transparentes, mas também uma mudança na forma como o aprendizado é avaliado, priorizando o processo de construção do saber em ambientes de inteligência híbrida (Círico, 2024), humana e tecnologias.

No campo da metodologia científica, Machado et al. (2024) propõem uma distinção operacional relevante para o debate sobre o uso ético e responsável, que preserva a importância do espaço da criatividade e do agenciamento epistêmico humano. Afirma que há uma diferença substantiva entre o uso instrumental da IA (na formatação, tradução, revisão gramatical e recuperação de informações) e o uso substitutivo, em que a ferramenta ocupa o lugar do pesquisador na formulação de problemas, na análise de dados e na interpretação de

resultados. O primeiro pode ser legítimo e produtivo, porém o segundo representa uma abdicação da autoria intelectual necessária. Os autores são categóricos na defesa da centralidade do elemento humano:

Destacamos, a priori, a primeira condição ética para o método que está em gênese, a qualidade valorativa dos dados depende do juízo da razão humana para a validação dos resultados e continuidade da pesquisa. Assim, a inteligência artificial não se torna uma substituta do labor humano. (Machado et al., 2024, p. 103).

Os autores afirmam que a inteligência natural e humana deve permanecer como a regente dos processos, atuando na condução dos comandos criativos e garantindo que a tecnologia seja apenas um recurso de agilidade, logo não deve ser um substituto do pensamento crítico e da agência do pesquisador. Dentro dessa perspectiva híbrida, a qualidade valorativa dos dados e a validação dos resultados dependem exclusivamente do juízo da razão humana, cabendo ao pesquisador orientar o foco da investigação e indicar os rumos das repetições realizadas pela ferramenta tecnológica. Dessa forma, a responsabilidade e a criatividade pelos processos permanecem com os usuários humanos, cujas capacidades focadas na ética e na razão servem como o contrapeso necessário à velocidade de processamento da máquina. A implementação dessa metodologia busca não apenas ganhos de performance e assertividade, mas também uma evolução positiva na literacia acadêmica, zelando sempre pelos pilares ético-científicos

que asseguram a transparência e a integridade da pesquisa científica.

Convém ressaltar que a criatividade científica não é um talento inato, tampouco é um processo misterioso inacessível à análise racional e ao processo educacional. Os pensamentos intuitivos e criativos de grandes pesquisadores resultam de um labor e domínio profundo sobre campos de estudos, de reiteradas leituras sistemáticas e pilares da literatura especializada, de diálogo e debates contínuos com pares, de engajamento persistente com problemas reais e de disposição para tolerar a incerteza, divergências, inquietações e dúvidas, que precedem qualquer descoberta ou análise genuína. Nenhuma dessas condições pode ser simulada por um simulador sistema treinado para produzir o texto estatisticamente mais mimético, provável e plausível. Essa concepção reafirma a insubstituibilidade e a indispensabilidade do pesquisador humano nas etapas que mais importam ao avanço do conhecimento.

7. O PESQUISADOR COMO CURADOR DO CONHECIMENTO: UMA IMPORTANTE COMPETÊNCIA EPISTÊMICA

A proliferação de ferramentas de IA generativa no ambiente universitário não torna o pesquisador obsoleto e ineficiente. Pelo contrário, este contexto cobra do cientista seu papel epistêmico mais central e mais exigente, na medida em que a abundância de *outputs* produzidos por sistemas não verificáveis exige um interlocutor humano capaz de distinguir o plausível do verdadeiro ou o mediano do inovador. O que emerge dessa nova conjuntura é uma competência que sempre existiu nas Universidades, quando os orientadores de pós-graduações se debruçam sobre as produções

dos seus orientandos, uma leitura epistêmica rigorosa e comprometida com a Ciência. No contexto do avanço das IAs no ambiente da pesquisa científica, essa curadoria crítica do conhecimento deve ser exercida por todos os usuários no processo de pesquisa auxiliado por IA generativas.

O termo curadoria, tal como empregado neste artigo, requer delimitação, uma vez que circula em pelo menos dois campos disciplinares com acepções técnicas distintas. Na Ciência da Informação, curadoria digital designa o conjunto de processos de seleção, organização, preservação e garantia de acesso contínuo a recursos informacionais ao longo do tempo, com ênfase na gestão do ciclo de vida dos dados e documentos (Sayão; Sales, 2012). Na Comunicação e no Jornalismo digital, o conceito migrou para designar a prática de agregar, filtrar e redistribuir conteúdos produzidos por terceiros em ambientes em rede, frequentemente com acréscimo de comentário ou enquadramento editorial. Nenhuma dessas acepções, tomadas isoladamente, abrange o que se propõe neste trabalho. A curadoria do conhecimento, no sentido aqui adotado, refere-se ao exercício deliberado e sistemático do **juízo epistêmico do pesquisador**, assim como os orientadores universitários, sobre os *outputs* gerados por sistemas de inteligência artificial generativa. Sua função volta-se para: julgar, criticar, avaliar, verificar, contextualizar, refutar ou validar cada produto do sistema antes de incorporá-lo ao processo de investigação. Trata-se de uma competência metodológica que opera sobre as fontes primárias da área do conhecimento, mas também sobre os resultados de um intermediário computacional cuja confiabilidade é estruturalmente variável.

Curadoria, nesse contexto, vai além de selecionar e organizar fontes, dimensão que já era constitutiva da prática de pesquisa e da Universidade. Curar significa julgar, selecionar bases e fontes de pesquisa significativas; encontrar lacunas de pesquisa para futuras e necessárias investigações e inovadoras; formular as questões investigativas que a IA não pode elaborar por si mesma; avaliar a qualidade, a precisão e a pertinência dos *outputs* gerados que possam contribuir com a pesquisa; conhecer as limitações estruturais dos sistemas de LLMs, suas funcionalidades e seus possíveis equívocos (Alucinações, Presentismo, Confabulações etc.); verificar toda citação e referência em bases de dados primárias (em artigos, livros, dissertações e teses); identificar os vieses algorítmicos e as lacunas sistemáticas de cada modelo usado; e, acima de tudo, assumir plena responsabilidade intelectual e ética pelo produto final da investigação. Nenhuma dessas operações avaliativas pode ser delegada ao *software* sem que o pesquisador deixe de ser pesquisador.

Tedesco e Ferreira (2023) situam esse imperativo de formação ética no campo dos fundamentos filosóficos, propondo que o enfrentamento dos desafios trazidos pela IA exige mais do que normas e regulamentos, requerendo a constituição de um *ethos*:

se quisermos avançar com as pesquisas em Educação, é preciso constituirmos, em nossos espaços acadêmicos, aquilo que a tradição filosófica nos legou como sabedoria prática [phronesis]. [...] Os estudantes e os docentes podem não se opor ao uso da IA, especialmente ao ChatGPT, desde que cultivem o ethos da sabedoria prática (phronesis)." (Tedesco; Ferreira, 2023, p. 18, 21).

Os autores (2023), afirmando “Ética e integridade acadêmica na pós-graduação em educação em tempos de inteligência artificial”, sustentam que o enfrentamento dos desafios trazidos pela Inteligência Artificial exige a constituição de um ethos da sabedoria prática (phronesis) nos espaços acadêmicos. Essa formação ética transcende o simples domínio técnico das ferramentas, caracterizando-se por uma reflexão crítica e filosófica sobre os limites, as implicações éticas e os riscos de uso inadequado da tecnologia, como a geração de textos falsos e o plágio. Nesse contexto, a competência essencial do pesquisador fundamenta-se na capacidade de análise crítica e criativa, habilidades que a IA não possui, sendo o sistema incapaz de avaliar a originalidade ou a qualidade de um argumento de forma autônoma. Portanto, a integridade acadêmica é compreendida como uma “sabedoria prática” e não legalista, que se manifesta no compromisso pessoal e no discernimento constante do pesquisador. A noção de phronesis, recuperada da tradição filosófica aristotélica, aponta para algo que nenhuma regulação normativa pode substituir. A disposição ativa de julgar bem em contextos singulares articula-se com as especificidades da situação concreta. São essas competências que o

pesquisador-curador precisa cultivar, exercer e se educar diante de quaisquer *outputs* produzidos por tecnologias.

O letramento epistêmico digital em IA (isto é, a compreensão dos mecanismos, das limitações e das implicações éticas dessas ferramentas) constitui atualmente uma competência transversal para todo pesquisador. A competência em questão é menos de domínio técnico especializado e mais de literacia epistêmica. O pesquisador-curador define-se menos por saber operar o ChatGPT e mais por saber como e quando não recorrer à ferramenta, e por quê. Esse discernimento é o que mais importa para a integridade do trabalho científico.

O quadro regulatório revela, com alguma ironia, que as próprias instituições universitárias têm chegado tarde a essa discussão. Cardoso e Gammarano (2025), em guia publicado na Revista de Ciências da Administração da UFSC, argumentam que o cenário acadêmico contemporâneo apresenta uma “lacuna crítica”. Há a necessidade de orientações éticas claras diante da adoção crescente de tecnologias generativas. O documento busca preencher esse vácuo fornecendo um referencial para o uso da IA Generativa (GenIA) em todas as etapas da pesquisa científica, da concepção teórica à redação final, visando a equilibrar a inovação tecnológica com a “responsabilidade epistêmica”. Para gerenciar o que denominam “Zonas de Risco Ético e Epistemológico”, os autores estabelecem critérios para diferenciar a assistência técnica da máquina da autoria humana, defendida como o núcleo inegociável da produção acadêmica. A sistematização dessas diretrizes fundamenta-se na análise de políticas editoriais e códigos de integridade nacionais e internacionais (como os da ANPAD, UNESCO

e Elsevier) que emergiram para consolidar um modelo de governança ética na ciência.

No âmbito das políticas institucionais para IA nas universidades, Chan (2023), em estudo publicado no *International Journal of Educational Technology in Higher Education*, propõe o *AI Ecological Education Policy Framework* (Estrutura de Políticas de Educação Ecológica em IA (ou Quadro de Políticas de Educação Ecológica em IA)). O quadro foi desenvolvido a partir de uma análise quantitativa e qualitativa com 637 participantes (457 estudantes de graduação e pós-graduação e 180 docentes e funcionários de universidades em Hong Kong), oferecendo parâmetros para que as instituições articulem a integração da IA em três dimensões fundamentais: 01 - Dimensão Pedagógica: Liderada pelos professores, focada na melhoria dos resultados de aprendizagem através da reformulação de avaliações e exames, no desenvolvimento de competências holísticas e na preparação dos estudantes para um mercado de trabalho impulsionado pela IA. 02 - Dimensão de Governança: Sob responsabilidade da alta gestão, estabelece diretrizes para a integridade acadêmica (prevenção de má conduta e plágio), privacidade de dados, transparência, segurança e garantia de equidade no acesso às tecnologias. 03 - Dimensão Operacional: Conduzida pelas equipes de TI e suporte ao ensino, focada na infraestrutura necessária, no monitoramento da implementação e no fornecimento de treinamento e suporte para a literacia em IA de estudantes e funcionários.

Em diretrizes internacionais, a UNESCO (2024) publicou os Quadros de Competências em IA para Estudantes e Professores, estruturado em quatro dimensões: mentalidade centrada no ser humano; ética da IA; técnicas e aplicações de IA; e concepção de sistemas de IA.

Para os pesquisadores universitários, o documento fornece referencial para a autoavaliação do grau de literacia epistêmica em IA, componente essencial do pesquisador-curador proposto neste artigo (UNESCO, 2024).

Trindade e Oliveira (2024) reforçam a necessidade de marcos regulatórios claros ao recuperar a posição da comunidade editorial internacional, destacando que os periódicos da Springer Nature estabeleceram dois princípios fundamentais: ferramentas de IA não podem ser listadas como autoras de pesquisas e os pesquisadores precisam documentar, obrigatoriamente, o uso dessas ferramentas na seção de métodos ou nos agradecimentos. Essa postura justifica-se pelo fato de a inteligência artificial não possuir características e habilidades humanas únicas, como experiência, capacidade de julgamento, responsabilidade e personalidade, o que a impede de assumir o ônus pela autoria de um trabalho científico. As autoras registram ainda a exigência mais ampla de que, à medida que os pesquisadores exploram as possibilidades dos *chatbots* avançados, os editores devem reconhecer seus usos legítimos, mas estabelecer diretrizes rígidas para evitar abusos. Entre as principais preocupações éticas citadas estão o risco de plágio e a apropriação de conteúdo sintetizado pela IA como se fosse produção original do pesquisador, o que compromete a integridade e a confiabilidade da ciência. Além disso, as ferramentas de IA generativa apresentam limitações técnicas severas, como a tendência de gerar informações falsas e fabricar referências bibliográficas ou identificadores DOI que não existem na realidade, o que exige um olhar rigorosamente crítico do autor humano. Diante desse cenário disruptivo, o desenvolvimento da competência em informação torna-se essencial, permitindo que os acadêmicos utilizem essas tecnologias de forma estratégica para tarefas como resumir literatura ou refinar a escrita, sem abrir mão do pensamento crítico e da inovação inerente ao

conhecimento científico. Portanto, a transparência e a aplicação de programas de letramento informacional são apontadas como os caminhos necessários para garantir que o uso da IA na academia ocorra de maneira ética, legal e responsável. O artigo afirma a necessidade de diretrizes:

à medida que os pesquisadores mergulham no admirável novo mundo dos avançados chatbots de IA, os editores precisam reconhecer seus usos legítimos e estabelecer diretrizes claras para evitar abusos. (Trindade; Oliveira, 2024, p. 09-10, apud Nature, 2023, tradução nossa).

No contexto brasileiro, Wotckoski et al. (2025), em revisão sistemática publicada na Revista DELOS, oferecem um argumento metodologicamente relevante ao afirmar que a solução para os problemas de integridade associados ao uso de IA não será resolvida por proibições ou soluções puramente tecnológicas. A iniciativa deve ser uma ação humana coordenada que articule políticas institucionais claras com uma profunda transformação pedagógica. Os autores sustentam que os detectores automáticos de texto são falíveis e representam um desvio do foco educacional, uma vez que a crise de integridade é um sintoma da natureza não neutra e opaca da tecnologia, cuja autonomia aparente frequentemente obscurece a responsabilidade estritamente humana. Essa opacidade, encapsulada no dilema da “caixa-preta”, dificulta ou impossibilita a compreensão de como o sistema chegou a uma decisão específica, o que viola a autonomia e o consentimento informado dos indivíduos afetados. Como a IA carece de agência moral e não pode

ser considerada uma entidade consciente, sendo um sistema probabilístico que mimetiza informações humanas, ela inevitavelmente aprende, reproduz e amplifica preconceitos históricos de gênero, raça e classe presentes em seus dados de treinamento, resultando em injustiças algorítmicas que marginalizam a diversidade e podem excluir sistematicamente grupos como pessoas com deficiência (Wotckoski et al., 2025).

A transformação pedagógica proposta (Wotckoski et al., 2025) para enfrentar esse cenário exige que a universidade reafirme sua missão de desenvolver o pensamento crítico, reconhecendo que o ato de escrever é, fundamentalmente, um ato de pensar. O uso da IAG como um atalho tomado pode arriscar a subverter o amadurecimento intelectual, que o ambiente acadêmico deve cultivar sempre. Isso implica uma modificação das práticas avaliativas tradicionais, em favor de avaliações que valorizem o processo de pesquisa, a argumentação oral e a resolução de problemas complexos. Complementarmente, a ação institucional deve ser pautada por uma governança proativa que estabeleça comitês de ética em IA para guiar o uso responsável da tecnologia e traduzir princípios abstratos em regras práticas, evitando o que a literatura chama de “ética de fachada”. O objetivo central reside no letramento digital e ético do corpo discente, capacitando o pesquisador como um agente epistêmico crítico, garantindo que a tecnologia sirva para potencializar a inteligência humana sem substituir a responsabilidade e o rigor intelectual que definem a missão do ensino superior.

“Diretrizes para o uso ético e responsável da Inteligência Artificial Generativa: um guia prático para pesquisadores”, estabelecidas por Sampaio, Sabbatini e Limongi (2024), aprofundam a necessidade de

uma transparência metodológica que não se limita apenas à disponibilização de dados, mas também exige a descrição detalhada de metadados da ferramenta, como nome, modelo, versão e a data exata da interação, já que esses sistemas são frequentemente atualizados e podem gerar resultados distintos ao longo do tempo. Essa transparência deve ser formalizada por meio de uma declaração específica no manuscrito, preferencialmente antes das referências bibliográficas, onde os autores explicitam como a tecnologia auxiliou na pesquisa e confirmam que o conteúdo foi revisado e editado sob rigor científico. Ao fornecer o log completo das interações e os scripts gerados, o pesquisador permite que a comunidade acadêmica audite o que os autores classificam como uma “caixa-preta”, garantindo assim a replicabilidade e a confiabilidade de processos.

No que tange à preservação da agência humana, o guia enfatiza o conceito de “humano no circuito” (*human in the loop*), posicionando o pesquisador como o único ente capaz de assumir a responsabilidade moral e legal pela integridade do trabalho. A curadoria mencionada deixa de ser um princípio abstrato para se tornar uma barreira crítica contra as “alucinações” e vieses intrínsecos aos modelos probabilísticos, que podem inventar fatos ou referências inexistentes. Portanto, o pesquisador deve atuar como supervisor e nunca como subordinado da máquina, garantindo que a Inteligência Artificial funcione estritamente como um assistente de pesquisa que potencializa a análise humana sem substituir o pensamento original, a criatividade e a independência intelectual. Essa postura assegura que o uso da tecnologia fortaleça o papel central do cientista, na produção de conhecimento soberano e eticamente orientado. Os autores são explícitos quanto ao que essa transparência exige na prática:

Pesquisadores que fizerem uso de Inteligência Artificial Generativa devem descrever na coverletter e no manuscrito como utilizaram a ferramenta para garantir transparência, replicabilidade e confiabilidade da pesquisa. [...] Como norma geral, isso deve incluir a disponibilização dos prompts e dos resultados advindos (outputs) do modelo, como os textos, os scripts de programação, visualizações e mesmo resultados de análise. (Sampaio; Sabbatini; Limongi, 2024, p. 20).

Mais do que uma formalidade burocrática, essa documentação integral é a prova concreta de que o pesquisador permaneceu no comando do processo, a materialização da curadoria enquanto sabedoria prática cotidiana e não apenas como princípio abstrato e legal.

Ante a precariedade ainda vigente dos marcos normativos, o cenário exige que o pesquisador desenvolva e exerça sempre sua *phronesis* (sabedoria prática), suas habilidades no uso responsável e ético de ferramentas tecnológicas, mas também desenvolva consciente e constantemente o seu letramento em IA para o avanço da Ciência e assuma inteira responsabilidade e autonomia, no processo de construção da pesquisa acadêmica. O pesquisador deve ser o curador do conhecimento e de si.

8. USO RESPONSÁVEL DA IA COMO FERRAMENTA INSTRUCIONAL NA PESQUISA UNIVERSITÁRIA

8.1. Princípios Orientadores

Alguns princípios têm respaldo em uma corrente consolidada de discussão filosófica sobre a ética da IA. Floridi et al. (2018), em artigo fundador do projeto AI4People publicado na *Minds and Machines*, propuseram um quadro ético estruturado em cinco princípios: beneficência, não maleficência, autonomia, justiça e explicabilidade. Esse quadro fundamenta a perspectiva de que o uso responsável da IA é uma decisão ética de caráter político e filosófico, influenciando diretamente os documentos normativos internacionais subsequentes (Floridi et al., 2018).

A proposta de Floridi et al. (2018) fundamenta-se na premissa de que a inteligência artificial não deve ser vista apenas como uma utilidade a ser regulada após sua maturidade, mas como uma força poderosa de “agência inteligente” que já está remodelando profundamente a vida humana e as interações sociais. Essa abordagem sustenta que a adoção de um quadro ético oferece uma “vantagem dupla” para as organizações e para a sociedade. Por um lado, permite identificar e aproveitar oportunidades de valor social que sejam aceitáveis, e, por outro, funciona como um sistema de alerta precoce para antecipar e evitar erros custosos ou cursos de ação que seriam rejeitados socialmente. O diferencial crucial desse quadro é a introdução da explicabilidade como um quinto princípio essencial, concebido para preencher a lacuna que os modelos éticos tradicionais deixam ao lidar com a IA. Este princípio é subdividido em duas dimensões: a inteligibilidade, que responde ao funcionamento técnico do sistema, e a responsabilidade (accountability), que determina quem deve prestar contas pelos resultados gerados. O objetivo final dessa estrutura é a construção de uma “Boa Sociedade de IA”, onde a tecnologia sirva para potencializar o florescimento humano em quatro frentes fundamentais: quem podemos nos tornar (autorrealização), o que podemos fazer (agência humana), o que

podemos alcançar (capacidades sociais) e como interagimos (coesão social) uns com os outros (Floridi et al., 2018).

No plano multilateral, a Recomendação sobre a Ética da Inteligência Artificial, adotada pela UNESCO em 2021 com adesão de 193 estados-membros, constitui o instrumento de maior alcance normativo sobre o tema. O documento articula dez princípios orientadores, entre os quais se destacam, para os fins deste artigo: proporcionalidade, proteção de dados e privacidade, supervisão humana efetiva, responsabilização. A Recomendação afirma que os sistemas de IA devem apoiar a capacidade de pesquisa humana, sem substituir o julgamento científico (UNESCO, 2021). Ainda, os Princípios sobre IA da OCDE (OECD, 2019), adotados pelo G20, estabeleceram cinco valores para o desenvolvimento responsável da IA: crescimento inclusivo e sustentável; respeito pelo Estado de direito e pelos direitos humanos; transparência e explicabilidade; robustez, segurança e proteção; e responsabilização. Esses princípios constituem a base sobre a qual o AI Act europeu e o PL 2.338/2023 brasileiro foram estruturados (OECD, 2019).

O uso responsável da IA na pesquisa universitária consiste em construir com a tecnologia uma relação metodologicamente informada e eticamente orientada. Nessa relação, o pesquisador define os termos do uso e preserva integralmente suas responsabilidades intelectuais. Com base na literatura revisada, é possível identificar seis princípios que concretizam e afirmam a necessidade ética e responsável dessa postura crítica.

01 - O primeiro princípio é a **transparência**. Todo uso de IA generativa em processos de pesquisa deve ser declarado explicitamente no manuscrito, com especificação da ferramenta,

versão, data de consulta e natureza do uso (Sampaio; Sabbatini; Limongi, 2024; Hayashi, 2025). Esse princípio ultrapassa a formalidade burocrática, pois alinha-se às recomendações do ICMJE, do COPE e das principais editoras científicas internacionais e constitui condição de replicabilidade da pesquisa. A Portaria CNPq n. 2.664/2026 eleva-o ao plano normativo vinculante ao definir integridade como:

Integridade: conduta que garante a qualidade e a idoneidade das informações em todos os aspectos do projeto de ciência, tecnologia e inovação desenvolvido, a partir da articulação de rígidos princípios éticos e morais, desde a concepção do estudo, até a finalização, publicação e comunicação pública dos resultados. (CNPq, 2026, Portaria n. 2.664, art. 5, XIV).

02 - A **verificação sistemática** constante é outro princípio. Toda informação, citação, referência ou dado produzido por IA deve ser conferido em fontes primárias, antes de ser incorporado ao trabalho (Lemos, 2024; Picalho; Oliveira; Cattivelli, 2025). A velocidade e a fluência da IA não medem acurácia, pois são características probabilísticas de sistemas projetados para gerar texto plausível, ainda que não necessariamente verdadeiro e factuais.

03 - O terceiro princípio, talvez o mais fundamental porque delimita a natureza do uso legítimo da IA, é o da **agência humana**. A autoria intelectual da pesquisa (a formulação do problema, o design metodológico, a análise dos dados, a interpretação dos resultados e

as conclusões) deve permanecer integralmente sob responsabilidade do pesquisador (Machado et al., 2024; Sampaio; Sabbatini; Limongi, 2024). A IA pode apoiar etapas operacionais do processo, mas não deve substituir o julgamento científico. Na seção "VI. Preservação da agência humana" do guia prático, há a observação:

A criação de conhecimento envolve um processo rigoroso que pode incluir teorização, teste de hipóteses, coleta de dados, análise, interpretação e avaliação cuidadosa, algo que a IA Generativa sozinha ainda não pode replicar. (Sampaio; Sabbatini; Limongi, 2024, p. 24)

04 - O princípio quarto é **o letramento epistêmico digital em IAG.**

O pesquisador precisa compreender os fundamentos do funcionamento dos LLMs (suas limitações probabilísticas, seus vieses de treinamento, o fenômeno das alucinações) para utilizá-los de forma genuinamente crítica (Tedesco; Ferreira, 2023; Bender et al., 2021). O que está em jogo é a literacia epistêmica, ou seja, saber o que, como e quando a ferramenta pode e, sobretudo, o que ela não pode fazer.

05 - O quinto princípio é o da **proporcionalidade entre ferramenta e agência humana.** Tarefas operacionais (revisão gramatical, formatação, tradução, sistematização de dados já coletados) são domínios nos quais a IA pode ser utilizada de forma legítima e produtiva. Já as tarefas epistêmicas centrais (formulação de hipóteses, análise crítica da literatura, interpretação de resultados)

exigem predominância do julgamento humano (Cardoso; Gammarano, 2025). Confundir esses dois domínios é a origem de boa parte dos problemas de integridade acadêmica documentados na literatura. Azevedo e Santos (2025) lembram que essa distinção remete a valores científicos inegociáveis:

há valores da ciência que são inegociáveis, como autoria responsável e comprometida, transparência e rastreabilidade das fontes, autonomia intelectual, compromisso com o debate informado, entre outros. (Azevedo; Santos, 2025, p. 12).

06 - Por fim, o **princípio da proteção de dados** é fundamental. O pesquisador deve evitar inserir em ferramentas de IA generativa dados sensíveis (nomes de participantes, dados não publicados, informações confidenciais), já que que a maioria dessas plataformas utiliza os *inputs* dos usuários para aprimorar seus modelos (Sampaio; Sabbatini; Limongi, 2024). O guia nacional é explícito, quando afirma que os pesquisadores “devem evitar carregar trabalhos não publicados ou sensíveis em sistemas de IA online, a menos que haja garantias de que os dados não serão reutilizados” (Sampaio; Sabbatini; Limongi, 2024, p. 23). Esse risco é especialmente relevante em pesquisas que envolvem seres humanos, onde a Lei Geral de Proteção de Dados (LGPD - Lei nº 13.709/2018) impõe obrigações específicas de tratamento e proteção de dados pessoais.

8.2. Usos Legítimos e Produtivos da IA na Pesquisa Universitária

Dentro desse quadro de princípios expostos, a literatura especializada aponta usos concretos da IA que podem genuinamente potencializar a pesquisa científica, sem comprometer sua integridade. Santos e Aganette (2026), em estudo publicado na RDBCI da Unicamp, investigam a integração da inteligência artificial (IA) no fluxo de trabalho da pesquisa científica, especificamente por meio da criação de um agente computacional baseado em modelos de LLMs para apoiar a revisão sistemática da literatura. Diante do crescimento exponencial das publicações na Ciência da Informação, os autores propuseram uma arquitetura capaz de realizar a leitura e extração semi-automatizada de elementos estruturais de artigos científicos, como objetivos, lacunas de pesquisa, metodologias, resultados e perspectivas futuras. A metodologia adotada foi uma pesquisa aplicada e qualitativa, utilizando um estudo de caso instrumental para validar a eficácia da ferramenta desenvolvida com as bibliotecas CrewAI, langchain_openai (utilizando o modelo GPT-4) e PyPDFLoader. Os resultados da pesquisa demonstram que o agente de IA possui um alto potencial para identificar corretamente os componentes estruturais dos textos acadêmicos, produzindo resumos que, em grande parte, são compatíveis com a interpretação humana. Conforme Santos e Aganette (2026), a ferramenta mostrou-se particularmente eficiente na automação de etapas iniciais e repetitivas, como a triagem de documentos e a extração de metadados, o que pode otimizar significativamente o tempo e o esforço dos pesquisadores. Dessa forma, a principal conclusão dos autores é que, embora os agentes de IA representem um avanço metodológico promissor para a Ciência, sua aplicação não substitui a necessidade de supervisão humana qualificada. A mediação humana permanece indispensável para garantir o rigor epistemológico, a análise crítica dos achados e a validação das

interpretações científicas produzidas pela máquina. O estudo reforça que o uso da IA deve ser ancorado em práticas éticas e responsáveis, servindo como uma ferramenta de apoio estratégico que potencializa a capacidade do pesquisador sem comprometer a qualidade da produção científica. Conforme os autores:

O agente identificou corretamente os elementos estruturais dos artigos e produziu resumos compatíveis com interpretações humanas em grande parte das dimensões analisadas. No entanto, apresentou limitações em profundidade analítica e contextualização, reforçando a necessidade de mediação humana. (Santos; Aganette, 2026, Resumo).

A atuação do agente revelou-se descritiva e pouco aprofundada em aspectos analíticos e conceituais, especialmente na interpretação das implicações dos achados e na formulação de inferências mais complexas. Esses pontos demonstram que, embora a tecnologia ofereça suporte relevante ao trabalho do pesquisador, sua atuação ainda exige mediação humana qualificada, capaz de agregar densidade teórica, crítica e contextual à análise automatizada. (Santos; Aganette, 2026, p. 18)

O uso de ferramentas especializadas, como por exemplo Undermind, Consensus, Elicit e Research Rabbit (projetadas para contextos de pesquisa científica, com consulta restrita a artigos revisados por pares), representa uma alternativa

metodologicamente mais segura do que o uso de assistentes conversacionais generalistas para levantamentos bibliográficos (Santos et al., 2025). Essas ferramentas reduzem significativamente o risco de alucinações ao operar sobre bases de dados curadas, embora não o eliminem completamente.

De acordo com o letramento epistêmico em IA, é possível identificar que as aplicações mais produtivas da IA estão associadas ao apoio: em levantamentos bibliográficos, na busca, na redação e organização de informações e dados em fontes delimitadas (fichamento de dados), em traduções, em revisões gramaticais e no desenvolvimento de competências metodológicas. Esses são domínios em que a ferramenta atua como suporte epistêmico e instrumental para o processo de aprendizagem e da pesquisa. O prévio e necessário letramento informacional do pesquisador é importante para a sua atuação ética e responsável, sem o qual a IA pode induzir vieses, omissões e dependência crítica, gerando dependência e “preguiça cognitiva”.

8.3. O Contexto Brasileiro: Regulação em Construção (2023-2025)

No Brasil, o cenário regulatório era ainda incipiente ao final de 2024, expondo pesquisadores, docentes e estudantes a uma zona de insegurança ética e metodológica, que demandava respostas institucionais urgentes. O levantamento de Sampaio, Sabbatini e Limongi (2024) evidenciou que as principais instâncias de financiamento e regulação da pesquisa científica no país (MEC, MCTI, CAPES, CNPq e agências estaduais de fomento) ainda não haviam publicado diretrizes formais sobre o uso de IA generativa.

Entre as iniciativas mais avançadas do período, destaca-se a publicação pela SciELO de guia sobre o uso de ferramentas de IA em setembro de 2023, que proíbe creditar a IA como autora e torna obrigatória a declaração de uso. O grupo de pesquisa *Understanding Artificial Intelligence* (UAI), vinculado ao IEA-USP, mantinha repositório com 80 diretrizes publicadas por IES do Brasil e de outros países, sendo os EUA os mais avançados nessa produção regulatória (Jornal da USP, 2025).

Cardoso e Gammarano (2025) documentam que as diretrizes específicas para a pesquisa acadêmica só começaram a emergir de forma mais consistente a partir de 2024, quando a pressão da comunidade científica e dos órgãos editoriais internacionais tornou o vácuo normativo insustentável (Cardoso; Gammarano, 2025).

8.4. A Regulação em Aceleração: O Cenário de Março de 2026

O início de 2026 marcou uma inflexão significativa no processo regulatório das IES brasileiras, evidenciando que o debate saiu da esfera informal e passou a ser tratado como questão institucional estratégica. Segundo levantamento publicado pelo jornal O Globo, em 16 de março de 2026, quase metade das universidades federais brasileiras já dispõe de guias institucionais ou se encontra em processo formal de debate e elaboração de normas para o uso da IA, em atividades acadêmicas e de pesquisa (O Globo, 2026). Esse dado representa uma mudança expressiva em relação ao cenário de 2024, quando apenas pouco mais de 4% das IES pesquisadas possuíam posicionamento oficial sobre o tema (Jornal da USP, 2025).

Em março de 2026, a Folha de São Paulo reportou que três das mais importantes universidades públicas do país (USP, Unicamp e Unesp) estavam estruturando departamentos específicos para estabelecer

protocolos de IA, coordenar pesquisas e promover a formação de alunos, professores e funcionários (ADUFG/Folha de São Paulo, 2026). A Unesp foi pioneira no sistema paulista. Além de publicar resolução geral e portaria específica para a pós-graduação em 2025, finalizou, em março de 2026, um guia normativo para a graduação estruturado em três categorias textuais explícitas. Conforme relatado pela Folha de São Paulo (ADUFG/Folha de São Paulo, 2026), o documento organiza as orientações em: “O que você PODE fazer”, “O que você NUNCA pode fazer” e “O que você TALVEZ possa fazer”. Na categoria dos usos permitidos figuram, por exemplo, “usar IA para traduzir textos, parafrasear parágrafos e elaborar resumos”; entre as vedações expressas constam “submeter trabalhos gerados por IA, de forma integral ou parcial, como se fossem produções originais” e “usar IA em avaliações sem autorização do professor” (ADUFG/Folha de São Paulo, 2026).

Na Unicamp, o protocolo institucional de uso de IA estava em fase final de aprovação pelo Conselho Universitário, em março de 2026, paralelamente à criação do Curso de bacharelado em Inteligência Artificial e Ciências de Dados. O coordenador do recém-criado Centro de Referência em Tecnologias de Inteligência Artificial da Unicamp, Leonardo Tomazeli Duarte, declarou que o documento “trata do tema de uma forma mais ampla, dando baliza para como encarar a questão da autoria, do plágio” (ADUFG/Folha de São Paulo, 2026).

A USP criou, em janeiro de 2026, o Escritório de Inteligência Artificial e Transformação Digital, ligado diretamente ao gabinete da sua Reitoria, com missão de planejar, coordenar e orientar a transformação digital e o uso de IA na universidade. O Centro de Inteligência Artificial e Aprendizado de Máquina (CIAAM) da USP

inaugurou, em fevereiro de 2026, um dos maiores clusters computacionais da América Latina, com tecnologia B200, e mantém convênio com o Ministério da Justiça para o desenvolvimento do primeiro Guia de Uso Ético de IA do país (ADUFG/Folha de São Paulo, 2026). Várias universidades federais também avançaram. A Universidade Federal de São Paulo (Unifesp) determinou que pesquisas e trabalhos acadêmicos devem informar explicitamente o uso de IA generativa. A Universidade Federal da Bahia (UFBA) orienta que os alunos indiquem qual ferramenta foi utilizada, quais comandos foram enviados e como as respostas foram interpretadas. Diversas instituições públicas estão promovendo o debate sério e crítico.

O conjunto dessas iniciativas converge em torno de quatro princípios fundamentais: (1) transparência obrigatória (não basta declarar que se usou IA; é necessário especificar ferramentas, versões, modelos e reproduzir os próprios prompts); (2) vedação à submissão de trabalhos integralmente gerados por IA como produção própria, sob pena de enquadramento como fraude acadêmica ou plágio; (3) revisão humana obrigatória de qualquer conteúdo produzido com auxílio de IA antes de sua utilização acadêmica ou didática; (4) reorientação das práticas avaliativas (provas orais, apresentações e projetos práticos) diante da ineficácia comprovada dos detectores automáticos de IA (ADUFG/Folha de São Paulo, 2026; O Globo, 2026).

No plano federal, o Ministério da Educação avançou em 2026 com o Referencial para Desenvolvimento e Uso Responsáveis de Inteligência Artificial na Educação, que estabelece diretrizes para assegurar que a IA respeite a dignidade humana, proteja a privacidade de estudantes e educadores, amplifique a capacidade

pedagógica dos professores e combata discriminações algorítmicas (MEC, 2026).

O conjunto de movimentos simultâneos nos planos institucional e federal evidencia que a governança da IA nas universidades brasileiras deixou de ser uma questão futura para se tornar uma demanda presente e urgente. O pesquisador universitário contemporâneo trabalha, portanto, em um contexto em que as regras estão sendo construídas em tempo simultâneo, o que torna ainda mais imperativa a posse de fundamentos teóricos e éticos sólidos para navegar nesse cenário com autonomia intelectual e responsabilidade científica.

8.5. O Quadro Jurídico e Normativo Brasileiro: Legislação, Instrumentos e Lacunas

A regulação da IA no Brasil estrutura-se em três planos normativos articulados: (1) a legislação federal de base já em vigor; (2) o processo legislativo específico para o Marco Legal da IA, em tramitação no Congresso Nacional; e (3) as normativas setoriais do Ministério da Educação (MEC) e do Conselho Nacional de Educação (CNE), voltadas especificamente ao ensino e à pesquisa acadêmica.

8.5.1. A Legislação de Base Vigente

Na ausência de uma lei específica sobre IA já promulgada, o ordenamento jurídico brasileiro aplicável ao uso de IA, na pesquisa e na educação, apoia-se sobre três pilares legislativos em vigor, cada qual com alcance próprio sobre dimensões específicas do problema.

O primeiro é a Lei Geral de Proteção de Dados Pessoais (LGPD), Lei nº 13.709, de 14 de agosto de 2018, que regula o tratamento de dados

peçoais por pessoas físicas e jurídicas, incluindo entes públicos, em todo o território nacional. A LGPD é diretamente pertinente ao uso de IA, na pesquisa acadêmica, porque ferramentas de IA generativa frequentemente recebem dados de usuários que podem ser utilizados para retreinamento dos modelos. O pesquisador que insere dados sensíveis de pesquisa (incluindo informações de participantes humanos), em plataformas de IA generativa, pode estar violando os princípios de finalidade, adequação e necessidade previstos na LGPD (Brasil, 2018). A Autoridade Nacional de Proteção de Dados (ANPD) é o órgão regulador competente.

Na literatura jurídica brasileira, Bioni (2021) constitui referência central (Proteção de dados pessoais: a função e os limites do consentimento) sobre proteção de dados pessoais no contexto da LGPD. O autor analisa a função e os limites do consentimento, como base legal para o tratamento de dados, argumentando que o consentimento isolado não é suficiente para garantir proteção efetiva em contextos de assimetria informacional. Para o pesquisador universitário, isso possui implicação direta, pois, ao inserir dados de participantes de pesquisa em plataformas de IA generativa, não se pode presumir que os termos de serviço dessas plataformas constituem base legal adequada sob a LGPD. É necessário verificar a política de tratamento de dados de cada plataforma e, quando aplicável, obter o consentimento específico dos participantes, especialmente em pesquisas que envolvam dados sensíveis ou populações vulneráveis (Bioni, 2021).

O segundo pilar é a Lei nº 12.965, de 23 de abril de 2014 (o Marco Civil da Internet), que estabelece princípios, garantias, direitos e deveres para o uso da internet no Brasil. Embora anterior à IA generativa, o Marco Civil fundamenta princípios de neutralidade, privacidade e

proteção de dados que informam a análise jurídica do uso de IA em ambientes digitais acadêmicos (Brasil, 2014).

O terceiro pilar é a Lei nº 14.533, de 11 de janeiro de 2023 (a Política Nacional de Educação Digital, PNED), que institui o letramento digital, a inclusão digital, a pesquisa e o desenvolvimento de tecnologias educacionais como eixos da política educacional brasileira. A PNED é o principal fundamento legal para as iniciativas do MEC em educação digital, incluindo as políticas de IA. O PL 2.338/2023 propõe alterar a PNED para incluir explicitamente competências para o uso de IA, entre elas o letramento algorítmico e o letramento crítico em IA (Brasil, 2023).

8.5.2. O PL 2.338/2023: O Marco Legal da IA em Tramitação

O instrumento legislativo central da regulação da IA no Brasil é o Projeto de Lei nº 2.338/2023, de autoria do então presidente do Senado Federal, senador mineiro Rodrigo Pacheco, na época. O PL “dispõe sobre o desenvolvimento, o fomento e o uso ético e responsável da inteligência artificial com base na centralidade da pessoa humana” (Brasil, 2023). Aprovado pelo Senado Federal em votação simbólica, em 10 de dezembro de 2024, o projeto passou a tramitar na Câmara dos Deputados sob relatoria do deputado Aguinaldo Ribeiro, vinculado a uma Comissão Especial presidida pela deputada Luísa Canziani (Senado Federal, 2024).

Em dezembro de 2025, o Poder Executivo encaminhou ao Congresso Nacional um projeto de lei complementar para corrigir um vício de iniciativa identificado no texto aprovado pelo Senado. O novo projeto cria o Sistema Nacional para Desenvolvimento, Regulação e Governança de Inteligência Artificial (SIA) e o Conselho

Brasileiro de Inteligência Artificial (CBIA), órgão estratégico de governança (Ministério da Gestão, 2025).

A votação final do PL 2.338/2023 na Câmara foi adiada para 2026 em razão de impasses políticos e divergências técnicas sobre o modelo regulatório. Em março de 2026, o texto ainda aguardava parecer do relator na Comissão Especial da Câmara (Câmara dos Deputados, 2026). Em termos de conteúdo, o PL organiza a regulação da IA por nível de risco, inspirado no modelo do AI Act europeu, aprovado pela União Europeia em 2024. As Diretrizes Éticas para uma IA Confiável, publicadas pelo Grupo de Alto Nível em Inteligência Artificial da Comissão Europeia (European Commission, 2019), definiram sete requisitos para sistemas de IA confiáveis: agência e supervisão humana, robustez técnica, privacidade e governança de dados, transparência, diversidade e não discriminação, bem-estar social e ambiental, e responsabilização.

8.5.3. O Referencial do MEC e as Diretrizes do CNE

No plano das políticas educacionais, o Ministério da Educação publicou, em março de 2026, o Referencial para Desenvolvimento e Uso Responsáveis de Inteligência Artificial na Educação, desenvolvido pela Secretaria de Gestão da Informação, Inovação e Avaliação de Políticas Educacionais (SEGAIPE). O Referencial estrutura suas orientações em torno de cinco diretrizes centrais, enunciadas nos seguintes termos:

Entre as diretrizes, destacam-se: a exigência de supervisão humana efetiva em processos educacionais que envolvam aplicações de inteligência artificial; o compromisso com a transparência e com a explicabilidade dos sistemas tecnológicos adotados; a observância rigorosa das leis que resguardam a proteção de dados pessoais e o direito à privacidade de estudantes e profissionais da educação; a valorização do trabalho docente e das práticas de formação inicial e continuada de professores; e o estímulo à constituição de ecossistemas de inovação abertos, colaborativos e comprometidos com o interesse público." (MEC, 2026).

O processo de construção do Referencial incluiu consulta pública realizada entre 10 e 29 de outubro de 2025, na plataforma Brasil Participativo (MEC, 2025), a primeira vez que o Estado brasileiro submeteu à escuta pública uma política nacional de IA para a educação.

Paralelamente, o Conselho Nacional de Educação (CNE) instalou, no início de 2025, a Comissão Bicameral para o Uso da Inteligência Artificial na Educação. A comissão realizou audiências públicas, ao longo de 2025, e o parecer final estava em fase de aprovação em março de 2026 (CNE, 2025). Os principais pontos identificados no parecer em elaboração, conforme documentado pela ANUP (2026), incluem: vinculação obrigatória do uso de IA a objetivos educacionais explícitos; supervisão permanente de profissionais da

educação; vedação da atuação pedagógica automatizada; e promoção do letramento digital, com ênfase na compreensão dos riscos, benefícios, princípios éticos e funcionamento básico dos modelos de IA.

Percebe-se em todas as legislações em desenvolvimento e em debates uma diretriz clara. A importância da implementação do letramento epistêmico digital em IA é urgente e necessário.

8.5.4. A Portaria CNPq N. 2.664/2026: Política de Integridade na Atividade Científica

A Portaria CNPq n. 2.664, assinada pelo presidente do Conselho Nacional de Desenvolvimento Científico e Tecnológico, é o marco normativo de alcance mais direto para a pesquisa científica financiada no Brasil, publicado em 06 de março de 2026 e publicada no Diário Oficial da União em 11 de março de 2026. O instrumento institui a Política de Integridade na Atividade Científica do CNPq, revogando a Portaria CNPq n. 1.735/2024 e a Resolução Normativa n. 006/2012 (CNPq, 2026).

O documento organiza-se em oito capítulos. Para os fins deste artigo, o art. 9, que estabelece as diretrizes de integridade na pesquisa, merece transcrição direta em seus dispositivos sobre IA generativa:

Art. 9. São diretrizes de integridade na pesquisa apoiada pelo CNPq: I - na atividade de pesquisa científica: [...] c) declarar o uso de ferramentas de Inteligência Artificial Generativa - IAG, de qualquer espécie e em qualquer fase do desenvolvimento da pesquisa (concepção, redação, análise de dados, submissão) especificando nos respectivos textos e exposições eletrônicas, a ferramenta utilizada e a finalidade; d) é vedada a submissão de conteúdo gerado por IAG como se fosse de autoria humana, sendo os autores integralmente responsáveis pelo conteúdo final, inclusive por eventuais plágios ou imprecisões geradas pela IAG; e) é vedada a inserção de projetos de pesquisa de terceiros em ferramentas de IAG para elaboração de pareceres científicos; f) responsabilizar-se integralmente pelo conteúdo final da pesquisa, inclusive por eventuais plágios ou imprecisões geradas pela IAG [...]. (CNPq, 2026, Portaria n. 2.664, art. 9, I, c-f).

No plano das infrações e sanções, a Portaria n. 2.664/2026 classifica como infração gravíssima a fabricação, falsificação ou manipulação fraudulenta de dados e a prática de plágio (art. 33, I-II), com sanções que podem incluir suspensão de bolsas e auxílios, impedimento para participação em ações de fomento por prazo determinado e suspensão do Currículo Lattes por período de três meses a um ano (art. 34, VIII). A Portaria sobre integridade afirma a centralidade da transparência, no uso ético e responsável das ferramentas de Inteligências Artificiais generativas.

A Portaria cria ainda a Comissão de Integridade na Atividade Científica (CIAC), órgão composto pelo Diretor Científico do CNPq e por quatro membros externos da comunidade científica, responsável pela promoção, prevenção, apuração e deliberação sobre desvios de integridade. As denúncias devem ser encaminhadas à Ouvidoria do CNPq por meio da Plataforma Fala.BR, assegurados o contraditório e a ampla defesa em todos os processos formais (art. 22-23).

Para o pesquisador universitário apoiado pelo CNPq, mas também para todos os estudiosos e pesquisadores brasileiros, a Portaria n. 2.664/2026 representa a primeira regulamentação normativa federal de caráter vinculante, que contempla obrigações específicas relativas ao uso de IA generativa, na pesquisa científica brasileira. Esses deveres reforçam o quadro de princípios articulado na seção 7.1 (7.1 Princípios orientadores) deste artigo, conferindo-lhes agora o peso normativo, que o cenário anterior não comportava. As Inteligências Artificiais podem ser instrumentos valiosos, no processo de pesquisa, mas devem ser apenas ferramentas. A autoria, o pensamento crítico, a responsabilidade, a inovação e a criatividade são inerentes à atuação do pesquisador-curador.

9. CONSIDERAÇÕES FINAIS

Ao longo deste artigo, buscou-se oferecer ao pesquisador universitário um mapa conceitual para navegar e compreender o fenômeno com a seriedade que ele exige e a criticidade que a formação científica impõe, mais do que um manual com diretrizes para o uso da IA. A análise empreendida permite afirmar que a IA generativa, em sua configuração atual, é simultaneamente uma tecnologia de amplo potencial operacional e de ameaça epistêmica

significativa, pois esse risco se materializa quando sua adoção ocorre sem o lastro crítico e reflexivo, que o ambiente universitário possui a obrigação de cultivar e promover.

A metáfora do “papagaio estocástico”, elaborada por Bender et al. (2021) e aliada ao “mímico mediano” para o contexto da pesquisa universitária, revela com precisão a natureza estrutural dos LLMs e o seu uso acrítico e irresponsável. São sistemas e usuários que reproduzem com fluência o conhecimento médio já produzido, mas sem compreensão semântica, sem capacidade de avaliação epistêmica e, fundamentalmente, sem acesso ao que ainda não foi dito e investigado. Nessa fronteira entre o já enunciado e o ainda por enunciar, o pesquisador-curador permanece insubstituível.

O fenômeno das alucinações, falhas e equívocos mapeados impõe, por sua vez, uma consequência metodológica não negociável no uso responsável e ético das ferramentas de Inteligências Artificiais. Toda informação gerada por IA deve ser verificada de forma independente com postura crítica e reflexiva, em fontes primárias. A fluência e a confiança retórica dos *outputs* não atestam verdade, pois os produtos gerados expressam propriedades estatísticas de sistemas treinados para produzir e veicular texto plausível. Aceitar esse texto sem verificação ultrapassa a simples negligência, já que reproduz, em escala acadêmica, o comportamento mimético que este artigo critica veementemente.

O pesquisador-curador emerge como a figura central de uma relação responsável com essas inovadoras ferramentas. O pesquisador responsável as utiliza como instrumentos poderosos, mas também falíveis como qualquer ferramenta. Além de leitor voraz, ele verifica sistematicamente os dados, mantendo intacta a

agência intelectual sobre cada decisão de pesquisa. Ao final, assume plena responsabilidade ética pela autoria do que assina. Assim, letramento em IA, transparência metodológica e verificação sistemática são as competências que definem essa postura crítica e que distinguem o uso legítimo da IA daquele que pode comprometer a integridade da Ciência.

O quadro normativo brasileiro está em construção acelerada (do Referencial do MEC e da Portaria do CNPq de 2026 ao parecer do CNE em votação e ao PL 2.338/2023 ainda em tramitação na Câmara). Entretanto, nenhum marco regulatório, por mais bem elaborado que seja, pode substituir a responsabilidade ética e a “sabedoria prática” que antecede e transcende a legislação. Para as universidades e para os seus educadores, o momento impõe grande urgência e responsabilidade. Como centros de pesquisas e produção do conhecimento, as universidades devem construir marcos institucionais claros e normativos, investir na formação e letramento epistêmico de docentes, servidores e pesquisadores, mas também deve participar ativamente da produção de conhecimento sobre as implicações dessas ferramentas para a investigação acadêmica e para a sociedade. A Ciência produzida nas universidades possui o dever e a capacidade de ser protagonista nessa construção.

Declaração de uso responsável de Inteligência Artificial: o presente artigo contou com apoio instrumental de assistentes de Inteligência Artificial generativa (Claude Sonnet 4.6, da Anthropic, e NotebookLM, do Google), empregados em funções estritamente auxiliares de organização estrutural, suporte à tradução, revisão textual e verificação de referências nas fontes consultadas. A formulação dos argumentos, a seleção e a verificação final das fontes, a análise crítica e a responsabilidade intelectual pelo

conteúdo são integralmente do autor, em conformidade com a Portaria CNPq n. 2.664/2026 e com as diretrizes COPE para integridade editorial.

REFERÊNCIAS BIBLIOGRÁFICAS

ADUFG-SINDICATO. **Universidades definem regras para o uso da IA.** Goiânia, 9 mar. 2026. Disponível em: <https://www.adufg.org.br/noticias/20-folha-de-sao-paulo/12515-universidades-definem-regras-para-o-uso-da-ia>. Acesso em: 17 mar. 2026.

ASSOCIAÇÃO NACIONAL DE UNIVERSIDADES PARTICULARES. **CNE deve votar em março parecer sobre uso de IA na educação básica e superior.** Brasília, DF, 25 fev. 2026. Disponível em: <https://anup.org.br/noticias/cne-deve-votar-em-marco-parecer-sobre-uso-de-ia-na-educacao-basica-e-superior/>. Acesso em: 17 mar. 2026.

AZAMBUJA, Celso Candido de; SILVA, Gabriel Ferreira da. Novos desafios para a educação na Era da Inteligência Artificial. **Filosofia Unisinos**, São Leopoldo, v. 25, n. 1, p. 1-16, 2024. DOI: 10.4013/fsu.2024.251.07.

AZEVEDO, Nathália Helena; SANTOS, Paulo Gabriel Franco dos. Primazia da dimensão utilitária e recuo crítico: inteligência artificial generativa e os valores em disputa na ciência. **Ensaio: Pesquisa em Educação em Ciências**, Belo Horizonte, v. 27, e59484, 2025. DOI: 10.1590/1983-2117-59484.

BENDER, Emily M.; GEBRU, Timnit; McMILLAN-MAJOR, Angelina; SHMITCHELL, Shmargaret. On the dangers of stochastic parrots: can

language models be too big? *In:* ACM CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY, 2021, Virtual event. **Proceedings** [...]. New York: Association for Computing Machinery, 2021. p. 610-623. DOI: 10.1145/3442188.3445922.

BHATTACHARYYA, Meहुl; MILLER, Valerie M.; BHATTACHARYYA, Debjani; MILLER, Larry E. High rates of fabricated and inaccurate references in ChatGPT-generated medical content. **Cureus**, [s. l.], v. 15, n. 5, e39238, 2023. DOI: 10.7759/cureus.39238.

BIONI, Bruno Ricardo. **Proteção de dados pessoais: a função e os limites do consentimento**. Rio de Janeiro: Forense, 2021.

BOURDIEU, Pierre. **Os usos sociais da ciência: por uma sociologia clínica do campo científico**. Tradução de Denice Barbara Catani. São Paulo: UNESP, 2004.

BRASIL. Câmara dos Deputados. **Projeto de Lei nº 2.338, de 2023**. Dispõe sobre o desenvolvimento, o fomento e o uso ético e responsável da inteligência artificial. Brasília, DF: Câmara dos Deputados, 2023. Disponível em: <https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2487262>. Acesso em: 17 mar. 2026.

BRASIL. Lei nº 12.965, de 23 de abril de 2014. Estabelece princípios, garantias, direitos e deveres para o uso da internet no Brasil. **Diário Oficial da União: seção 1**, Brasília, DF, 24 abr. 2014. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l12965.htm. Acesso em: 17 mar. 2026.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). **Diário Oficial da União: seção 1**, Brasília,

DF, 15 ago. 2018. Disponível em:
https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 17 mar. 2026.

BRASIL. Lei nº 14.533, de 11 de janeiro de 2023. Institui a Política Nacional de Educação Digital e altera as Leis nº 9.394, de 20 de dezembro de 1996, nº 9.448, de 14 de março de 1997, nº 10.260, de 12 de julho de 2001, e nº 10.753, de 30 de outubro de 2003. **Diário Oficial da União: seção 1**, Brasília, DF, 12 jan. 2023. Disponível em:
https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2023/lei/l14533.htm. Acesso em: 17 mar. 2026.

CARDOSO, Janaína Gularte; GAMMARANO, Igor de Jesus Lobato Pompeu. Guia para o uso ético e responsável da inteligência artificial generativa no âmbito acadêmico. **Revista de Ciências da Administração**, Florianópolis, v. 27, n. 67, p. 1-54, 2025. DOI: 10.5007/2175-8077.2025.e109536.

CHAN, Cecilia Ka Yuk. A comprehensive AI policy education framework for university teaching and learning. **International Journal of Educational Technology in Higher Education**, [s. l.], v. 20, art. 38, 2023. DOI: 10.1186/s41239-023-00408-3.

CÍRICO, Juh. Reflexões sobre ética no uso de Inteligência Artificial em pesquisas no campo da Educação Profissional e Tecnológica. **Revista Brasileira da Educação Profissional e Tecnológica**, Natal, v. 2, n. 24, e17376, 2024. DOI: 10.15628/rbept.2024.17376.

CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO. Portaria CNPq nº 2.664, de 6 de março de 2026. Institui a Política de Integridade na Atividade Científica do CNPq. **Diário Oficial da União: seção 1**, Brasília, DF, n. 47, p. 4, 11 mar. 2026.

Disponível em:
http://memoria2.cnpq.br/web/guest/view/-/journal_content/56_INSTANACE_0oED/10157/23142775?COMPANY_ID=10132. Acesso em: 9 abr. 2026.

CONSELHO NACIONAL DE EDUCAÇÃO. **Comissão Bicameral para o Uso da Inteligência Artificial na Educação**. Brasília, DF: CNE/MEC, 2025. Disponível em:
<https://brasilparticipativo.presidencia.gov.br/processes/iaeducacaobrasileira>. Acesso em: 9 abr. 2026.

DOSHI, Anil R.; HAUSER, Oliver P. Generative AI enhances individual creativity but reduces the collective diversity of novel content. **Science Advances**, Washington, DC, v. 10, n. 28, eadn5290, 2024. DOI: 10.1126/sciadv.adn5290.

EUROPEAN COMMISSION. **Ethics guidelines for trustworthy AI**. Brussels: High-Level Expert Group on Artificial Intelligence, 2019. Disponível em: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>. Acesso em: 9 abr. 2026.

FAN, Yizhou; TANG, Luzhen; LE, Huixiao; SHEN, Kejie; TAN, Shufang; ZHAO, Yueying; SHEN, Yuan; LI, Xinyu; GAŠEVIĆ, Dragan. Beware of metacognitive laziness: effects of generative artificial intelligence on learning motivation, processes, and performance. **British Journal of Educational Technology**, [s. l.], v. 56, n. 2, p. 489-530, 2025. DOI: 10.1111/bjet.13544.

FARQUHAR, Sebastian; KOSSEN, Jannik; KUHN, Lorenz; GAL, Yarin. Detecting hallucinations in large language models using semantic entropy. **Nature**, London, v. 630, n. 8017, p. 625-630, 2024. DOI: 10.1038/s41586-024-07421-0.

FLORIDI, Luciano; COWLS, Josh; BELTRAMETTI, Monica; CHATILA, Raja; CHAZERAND, Patrice; DIGNUM, Virginia; LUETGE, Christoph; MADELIN, Robert; PAGALLO, Ugo; ROSSI, Francesca; SCHAFER, Burkhard; VALCKE, Peggy; VAYENA, Effy. An ethical framework for a good AI society: opportunities, risks, principles, and recommendations. **Minds and Machines**, Dordrecht, v. 28, n. 4, p. 689-707, 2018. DOI: 10.1007/s11023-018-9482-5.

HAYASHI, Maria Cristina Piumbato Innocentini. Ética e Inteligência Artificial na comunicação científica: desafios para os periódicos científicos. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, v. 30, e103497, 2025. DOI: 10.5007/1518-2924.2025.e103497.

HICKS, Michael Townsen; HUMPHRIES, James; SLATER, Joe. ChatGPT is bullshit. **Ethics and Information Technology**, Dordrecht, v. 26, n. 2, art. 38, 2024. DOI: 10.1007/s10676-024-09775-5.

JI, Ziwei; LEE, Nayeon; FRIESKE, Rita; YU, Tiezheng; SU, Dan; XU, Yan; ISHII, Etsuko; BANG, Yejin; MADOTTO, Andrea; FUNG, Pascale. Survey of hallucination in natural language generation. **ACM Computing Surveys**, New York, v. 55, n. 12, art. 248, p. 1-38, 2023. DOI: 10.1145/3571730.

JORNAL DA USP. **Uma estrutura de governança de IA para instituições de ensino superior: riscos e desafios**. São Paulo: USP, 2025. Disponível em: <https://jornal.usp.br/artigos/uma-estrutura-de-governanca-de-ia-para-instituicoes-de-ensino-superior-riscos-e-desafios/>. Acesso em: 17 mar. 2026.

LEMOS, André Luiz Martins. Erros, falhas e perturbações digitais em alucinações das IA generativas: tipologia, premissas e epistemologia

da comunicação. **MATRIZES**, São Paulo, v. 18, n. 1, p. 75-91, 2024. DOI: 10.11606/issn.1982-8160.v18i1p75-91.

LIMA, Cleosanice Barbosa; SERRANO, Agostinho. Inteligência Artificial Generativa e ChatGPT: uma investigação sobre seu potencial na educação. **Transinformação**, Campinas, v. 36, e2410839, 2024. DOI: 10.1590/2318-0889202436e2410839.

MACHADO, Ariél Philippi; ANDRADE, Cleuciane Cezar de; MOREANO, Tatiane Ribeiro; RAUBER, Alexandre. Metodologia de pesquisa em uso de inteligência artificial generativa: reflexões éticas e científicas na prática acadêmica. **Annales FAJE**, Belo Horizonte, v. 9, n. 5, p. 100-107, 2024. Disponível em: <https://www.faje.edu.br/periodicos/index.php/annales/article/view/5946>. Acesso em: 27 maio 2026.

MARKEN, Stephanie. AI is routine for college students, despite campus limits. **Gallup**, Washington, DC, 2 abr. 2026. Disponível em: <https://news.gallup.com/poll/704090/routine-college-students-despite-campus-limits.aspx>. Acesso em: 25 maio 2026.

MESSERI, Lisa; CROCKETT, M. J. Artificial intelligence and illusions of understanding in scientific research. **Nature**, London, v. 627, n. 8002, p. 49-58, 2024. DOI: 10.1038/s41586-024-07146-0.

MINISTÉRIO DA EDUCAÇÃO. **MEC abrirá consulta pública sobre IA na educação**. Brasília, DF: MEC, 8 out. 2025. Disponível em: <https://www.gov.br/mec/pt-br/assuntos/noticias/2025/outubro/mec-abrira-consulta-publica-sobre-ia-na-educacao>. Acesso em: 17 mar. 2026.

MINISTÉRIO DA EDUCAÇÃO. **Referencial para desenvolvimento e uso responsáveis de inteligência artificial na educação.** Brasília, DF: SEGAPE/MEC, fev. 2026. Disponível em: <https://www.gov.br/mec/pt-br/media/segape/referencial-oficial-pt.pdf>. Acesso em: 30 mar. 2026.

MINISTÉRIO DA GESTÃO E DA INOVAÇÃO EM SERVIÇOS PÚBLICOS. **PL do governo propõe sistema de governança para a inteligência artificial no país.** Brasília, DF, 9 dez. 2025. Disponível em: <https://www.gov.br/gestao/pt-br/assuntos/noticias/2025/dezembro/pl-do-governo-propoe-sistema-de-governanca-para-a-inteligencia-artificial-no-pais>. Acesso em: 12 jun. 2026.

NAM, Jane. 56% of college students have used AI on assignments or exams. **BestColleges**, [s. l.], 22 nov. 2023. Disponível em: <https://www.bestcolleges.com/research/most-college-students-have-used-ai-survey/>. Acesso em: 28 maio 2026.

O GLOBO. **Limite aos chatbots: quase metade das universidades federais tem guias ou debate regras para usar IA.** Rio de Janeiro, 16 mar. 2026. Disponível em: <https://oglobo.globo.com/brasil/educacao/noticia/2026/03/16/limite-aos-chatbots-quase-metade-das-universidades-federais-tem-guias-ou-debate-regras-para-usar-ia.ghtml>. Acesso em: 28 mar. 2026.

OBSERVATÓRIO FUNDAÇÃO ITAÚ; DATAFOLHA. **Pesquisa sobre consumo e uso de inteligência artificial no Brasil.** São Paulo: Fundação Itaú, 2025. Disponível em: <https://fundacaoitau.org.br/observatorio/biblioteca/pesquisa-consumo-e-uso-da-inteligencia-artificial-no-brasil>. Acesso em: 25 maio 2026.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. **OECD principles on artificial intelligence.** Paris: OECD, 2019. Disponível em: <https://oecd.ai/en/ai-principles>. Acesso em: 9 abr. 2026.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS PARA A EDUCAÇÃO, A CIÊNCIA E A CULTURA. **AI competency frameworks for students and teachers.** Paris: UNESCO, 2024. Disponível em: <https://www.unesco.org/en/digital-education/ai-future-learning>. Acesso em: 9 abr. 2026.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS PARA A EDUCAÇÃO, A CIÊNCIA E A CULTURA. **Recommendation on the ethics of artificial intelligence.** Paris: UNESCO, 2021. Disponível em: <https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence>. Acesso em: 9 abr. 2026.

PETERS, Uwe; CHIN-YEE, Benjamin. Generalization bias in large language model summarization of scientific research. **Royal Society Open Science**, London, v. 12, n. 4, art. 241776, 2025. DOI: 10.1098/rsos.241776.

PICALHO, Antonio Carlos; OLIVEIRA, Gisele Rosa de; CATIVELLI, Adriana Stefani. Inteligência artificial no levantamento bibliográfico em bases de dados científicos: comparando expressões de busca no ChatGPT, Copilot e Gemini. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 23, e025013, 2025. DOI: 10.20396/rdbci.v23i00.8678378.

SAMPAIO, Rafael Cardoso; NICOLÁS, Maria Alejandra; JUNQUILHO, Tainá Aguiar; SILVA, Luiz Rogério Lopes; FREITAS, Christiana Soares de; TELLES, Márcio; TEIXEIRA, João Senna; ESCÓSSIA, Fernanda da;

SANTOS, Luiza Carolina dos. ChatGPT e outras IAs transformarão a pesquisa científica: reflexões sobre seus usos. **Revista de Sociologia e Política**, Curitiba, v. 32, e008, 2024. DOI: 10.1590/1678-98732432e008.

SAMPAIO, Rafael Cardoso; SABBATINI, Marcelo; LIMONGI, Ricardo. **Diretrizes para o uso ético e responsável da Inteligência Artificial Generativa**: um guia prático para pesquisadores. São Paulo: Intercom, 2024.

SAMPAIO, Rafael Cardoso. O paradoxo da transparência no uso de IA generativa na pesquisa acadêmica. SciELO em Perspectiva, São Paulo, 10 out. 2025. Disponível em: <https://blog.scielo.org/blog/2025/10/10/o-paradoxo-da-transparencia-no-uso-de-ia-generativa-na-pesquisa-academica/>. Acesso em: 17 mar. 2026.

SANTAELLA, Lucia; KAUFMAN, Dora. A Inteligência Artificial generativa como quarta ferida narcísica do humano. **MATRIZES**, São Paulo, v. 18, n. 1, p. 37-53, 2024. DOI: 10.11606/issn.1982-8160.v18i1p37-53.

SANTOS, Daiana Amado Baptista dos; CUNHA, Joice Soltosky; CARVALHO, Maria Luisa de; TEIXEIRA, Raquel da Silva. Inteligência artificial na pesquisa acadêmica: contribuições e experiências de bibliotecas acadêmicas. ***In:*** SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 23., 2025, São Paulo. Anais [...]. São Paulo: FEBAB, 2025.

SANTOS, Rodrigo Fernandes dos; AGANETTE, Elisangela Cristina. Inteligência artificial e automação na pesquisa científica: uma proposta de agente computacional para apoio à revisão da literatura. **RDBCI: Revista Digital de Biblioteconomia e Ciência da**

Informação, Campinas, v. 24, e026012, 2026. DOI: 10.20396/rdbci.v24i00.8679621.

SAYÃO, Luís Fernando; SALES, Luana Farias. Curadoria digital: um novo patamar para preservação de dados digitais de pesquisa. **Informação & Sociedade: Estudos**, João Pessoa, v. 22, n. 3, p. 179-191, set./dez. 2012.

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

SENADO FEDERAL. **Senado aprova marco regulatório da inteligência artificial (PL 2.338/2023)**. Brasília, DF, 10 dez. 2024. Disponível em: <https://www12.senado.leg.br/noticias/materias/2024/12/10/senado-aprova-regulamentacao-da-inteligencia-artificial-texto-vai-a-camara>. Acesso em: 17 mar. 2026.

SPINAK, Ernesto. Inteligência Artificial e a comunicação da pesquisa. **SciELO em Perspectiva**, São Paulo, 30 ago. 2023. Disponível em: <https://blog.scielo.org/blog/2023/08/30/inteligencia-artificial-e-a-comunicacao-da-pesquisa/>. Acesso em: 17 mar. 2026.

STEPHENSON, Rose; ARMSTRONG, Charlotte. **Student generative AI survey 2026**. Oxford: Higher Education Policy Institute; Kortext, mar. 2026. (HEPI Report, 199). ISBN 978-1-915744-59-3. Disponível em: <https://www.hepi.ac.uk/reports/student-generative-ai-survey-2026/>. Acesso em: 25 maio 2026.

TEDESCO, Anderson Luiz; FERREIRA, Jacques de Lima. Ética e integridade acadêmica na pós-graduação em educação em tempos

de inteligência artificial. **Periódico Horizontes**, São Paulo, v. 41, n. 1, e023032, 2023. DOI: 10.24933/horizontes.v41i1.1620.

TRINDADE, Alessandra Stefane Cândido Elias da; OLIVEIRA, Henry Poncio Cruz de. Inteligência Artificial (IA) generativa e competência em informação: habilidades informacionais necessárias ao uso de ferramentas de IA generativa em demandas informacionais de natureza acadêmica-científica. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 29, e-47485, 2024. DOI: 10.1590/1981-5344/47485.

WALTERS, William H.; WILDER, Esther I. Fabrication and errors in the bibliographic citations generated by ChatGPT. **Scientific Reports**, London, v. 13, n. 1, 14045, 2023. DOI: 10.1038/s41598-023-41032-5.

WOTCKOSKI, Ricardo Boone; BORGES, Marilurdes Cruz; CELESTE, Marcos; ANDRADE, Simone Tavares de; OLIVEIRA, Cláudia de Fátima. Inteligência artificial generativa e seus desafios éticos no ensino superior: uma revisão sistemática da literatura. **Revista DELOS**, Curitiba, v. 18, n. 70, e6187, 2025. DOI: 10.55905/rdelosv18.n70-061.

XU, Ziwei; JAIN, Sanjay; KANKANHALLI, Mohan. Hallucination is inevitable: an innate limitation of large language models. **arXiv**, [s. l.], arXiv:2401.11817, 2024. Disponível em: <https://arxiv.org/abs/2401.11817>. Acesso em: 17 mar. 2026.

YAN, Lixiang; GREIFF, Samuel; TEUBER, Ziwen; GAŠEVIĆ, Dragan. Promises and challenges of generative artificial intelligence for human learning. **Nature Human Behaviour**, London, v. 8, n. 10, p. 1839-1850, 2024. DOI: 10.1038/s41562-024-02004-5.

ZAWACKI-RICHTER, Olaf; MARÍN, Victoria I.; BOND, Melissa; GOUVERNEUR, Franziska. Systematic review of research on artificial intelligence applications in higher education: where are the educators? **International Journal of Educational Technology in Higher Education**, [s. l.], v. 16, art. 39, 2019. DOI: 10.1186/s41239-019-0171-0.

¹ Professor Titular do Departamento de Filosofia e Métodos (DFIME) da Universidade Federal de São João del Rei - UFSJ / Brasil Linhas de Ensino Superior: Docente de Metodologia e Comunicação Científicas / Literatura, Mitologia e Teoria Psicanalítica e Junguiana. Linhas de Pesquisa: Literaturas; Psicologia Analítica (Jung); Mitologias Comparadas. Linhas de Extensão: Literaturas; Psicologia Analítica (Jung) e Mitologias

² Possui graduação em Sistemas de Informação pela Universidade Presidente Antônio Carlos (2008), especialização em Engenharia de Sistemas pela Escola Superior Aberta do Brasil(2015), Mestrado Profissional em Tecnologia, Ambiente e Sociedade pela Universidade Federal dos Vales do Jequitinhonha e Mucuri(2018) e Doutorado em Organização Gestão do Conhecimento (PPGGOC) da Escola de Ciência da Informação Universidade Federal de Minas Gerais (UFMG) com período sanduíche na Universidade de Viena pelo PDSE - Programa de Doutorado Sanduíche no Exterior. Atualmente é técnico laboratório de informática da Universidade Federal dos Vales do Jequitinhonha e Mucuri - Campus Mucuri. Tem experiência na área de Tecnologia. ORCID: <https://orcid.org/0000-0001-8387-5272>. Lattes: <http://lattes.cnpq.br/1308905548372342>.