Informatio
31(2), 2026, e701 ISSN: 2301-1378

DOI: 10.35643/Info.31.2.17

Resenha

Da declaração à rastreabilidade: governança editorial da inteligência artificial na Ciência Aberta

Resenha: Vechi, B. D., Shintaku, M., Maia, M. A. Q., e Sousa, R. P. M. de. (2026). *Inteligência artificial na editoração científica em Ciência Aberta: riscos, integridade científica e governança editorial*.

From disclosure to traceability: editorial governance of artificial intelligence in Open Science

De la declaración a la trazabilidad: gobernanza editorial de la inteligencia artificial en la Ciencia Abierta

José Eduardo Santarem Segundo[a] ORCID: 0000-0003-3360-7872

[a]Universidade de São Paulo (USP), Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Departamento de Educação, Informação e Comunicação. Av. Bandeirantes, 3900, Monte Alegre, CEP 14040-901, Ribeirão Preto, SP, Brasil. E-mail: santarem@usp.br.

Resumo

Resenha do livro *Inteligência artificial na editoração científica em Ciência Aberta: riscos, integridade científica e governança editorial*, de Bernardo Dionízio Vechi, Milton Shintaku, Maria Aniolly Queiroz Maia e Rosilene Paiva Marinho de Sousa (Editora Ibict, 2026). A obra mapeia os pontos de entrada da inteligência artificial no fluxo editorial e converte princípios de integridade em rotinas verificáveis, organizadas por etapa (submissão, avaliação, edição de texto, curadoria de metadados e pós-publicação). A resenha descreve a estrutura do livro e o avalia a partir da perspectiva da confiabilidade informacional, da governança com IA e do uso de agentes de inteligência artificial. Destacam-se a centralidade atribuída aos metadados e à rastreabilidade, o tratamento das saídas automatizadas como sinais e não como veredictos e a atenção à injeção de instruções ocultas em manuscritos. Como limites e agenda de pesquisa, apontam-se a ausência de validação empírica do modelo, a pouca exploração de representações legíveis por máquina para declarações de uso de IA, as tensões entre sondas ocultas e acessibilidade e o papel dos periódicos como produtores dos corpus que alimentam sistemas generativos. Conclui-se que a obra é referência oportuna e de elevado valor prático para editores ibero-americanos.

Santarem Segundo, J.E. (2026). Da declaração à rastreabilidade: governança editorial da inteligência artificial na Ciência Aberta Resenha: Vechi, B. D., Shintaku, M., Maia, M. A. Q., e Sousa, R. P. M. de. (2026). Inteligência artificial na editoração científica em Ciência Aberta: riscos, integridade científica e governança editorial.. *Informatio*, 31(2), e701. https://doi.org/10.35643/Info.31.2.17

Palavras-chave: INTELIGÊNCIA ARTIFICIAL; EDITORAÇÃO CIENTÍFICA; GOVERNANÇA EDITORIAL; INTEGRIDADE CIENTÍFICA; CONFIABILIDADE INFORMACIONAL.

Abstract

Review of the book *Inteligência artificial na editoração científica em Ciência Aberta: riscos, integridade científica e governança editorial*, by Bernardo Dionízio Vechi, Milton Shintaku, Maria Aniolly Queiroz Maia and Rosilene Paiva Marinho de Sousa (Editora Ibict, 2026). The book maps the entry points of artificial intelligence into the editorial workflow and translates integrity principles into verifiable routines, organized by stage (submission, peer review, copyediting, metadata curation and post-publication). The review describes the structure of the book and assesses it from the perspective of information reliability, the semantic web and the use of AI agents. It highlights the central role given to metadata and traceability, the treatment of automated outputs as signals rather than verdicts, and the attention paid to hidden prompt injection in manuscripts. As limitations and agenda, it notes the lack of empirical validation of the model, the limited exploration of machine-readable representations of AI-use disclosures, the tensions between hidden probes and accessibility, and the role of journals as producers of the corpora that feed generative systems. It concludes that the book is a timely reference of high practical value for Ibero-American editors.

Keywords: ARTIFICIAL INTELLIGENCE; SCHOLARLY PUBLISHING; EDITORIAL GOVERNANCE; RESEARCH INTEGRITY; INFORMATION RELIABILITY.

Resumen

Reseña del libro *Inteligência artificial na editoração científica em Ciência Aberta: riscos, integridade científica e governança editorial*, de Bernardo Dionízio Vechi, Milton Shintaku, Maria Aniolly Queiroz Maia y Rosilene Paiva Marinho de Sousa (Editora Ibict, 2026). La obra mapea los puntos de entrada de la inteligencia artificial en el flujo editorial y convierte principios de integridad en rutinas verificables, organizadas por etapa (envío, evaluación, edición de texto, curaduría de metadatos y pospublicación). La reseña describe la estructura del libro y lo evalúa desde la perspectiva de la confiabilidad informacional, la web semántica y el uso de agentes de inteligencia artificial. Se destacan la centralidad atribuida a los metadatos y a la trazabilidad, el tratamiento de las salidas automatizadas como señales y no como veredictos, y la atención a la inyección de instrucciones ocultas en manuscritos. Como límites y agenda, se señalan la ausencia de validación empírica del modelo, la escasa exploración de representaciones legibles por máquina para las declaraciones de uso de IA, las tensiones entre sondas ocultas y accesibilidad, y el papel de las revistas como productoras de los corpus que alimentan sistemas generativos. Se concluye que es una referencia oportuna y de alto valor práctico para editores iberoamericanos.

Palabras clave: INTELIGENCIA ARTIFICIAL; EDICIÓN CIENTÍFICA; GOBERNANZA EDITORIAL; INTEGRIDAD CIENTÍFICA; CONFIABILIDAD INFORMACIONAL.

Apresentação

Poucas atividades acadêmicas foram tão rapidamente atravessadas pela inteligência artificial (IA) generativa quanto a editoração científica. Autores recorrem a modelos de linguagem para escrever, traduzir e revisar; avaliadores, nem sempre de forma declarada, para resumir manuscritos e redigir pareceres; e sistemas de gestão editorial incorporam rotinas automatizadas de triagem, similaridade e extração de metadados. É nesse cenário que se insere *Inteligência artificial na editoração científica em Ciência Aberta: riscos, integridade científica e governança editorial*, publicado em 2026 pela Editora do Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) como e-book de acesso aberto sob licença CC BY 4.0, com prefácio de Hugo Valadares Siqueira.

A autoria reúne perfis complementares: Bernardo Dionízio Vechi e Maria Aniolly Queiroz Maia, bibliotecários e pesquisadores do Ibict com atuação em editoração, metadados e no Open Journal Systems (OJS); Milton Shintaku, coordenador de Tecnologias para Informação do Instituto e autor de ampla produção sobre Ciência Aberta para editores; e Rosilene Paiva Marinho de Sousa, docente da Universidade Federal da Paraíba, com formação também em Direito e pesquisa em propriedade intelectual e proteção de dados. Essa composição se reflete no texto, que combina conhecimento do cotidiano dos periódicos, domínio da infraestrutura informacional e atenção aos aspectos jurídicos, da Lei Geral de Proteção de Dados (LGPD) às políticas autorais.

O objetivo declarado é “mapear os pontos de entrada da IA no fluxo editorial e analisar como sua adoção interfere na governança editorial em contextos orientados pela Ciência Aberta” (Vechi, Shintaku, Maia, e Sousa, 2026, p. 10). A tese que sustenta o livro é que a IA “não altera somente o ritmo e as rotinas do fazer editorial”, mas “promove uma mudança estrutural da governança” (p. 9), deslocando o problema da eficiência para os arranjos decisórios, a responsabilização e a rastreabilidade.

Leio a obra a partir de um lugar específico: o de pesquisador da Ciência da Informação dedicado à web semântica e à confiabilidade informacional de sistemas baseados em grandes modelos de linguagem (LLM) com geração aumentada por recuperação (*Retrieval-Augmented Generation*, RAG), e desenvolvedor e usuário cotidiano de agentes de IA. Esse ponto de vista orienta tanto o reconhecimento das contribuições quanto as questões que proponho ao final.

Estrutura e argumento

O livro, de 80 páginas, organiza-se em sete capítulos que avançam da contextualização para a prescrição. A introdução situa a reconfiguração da comunicação científica (expansão do acesso aberto, aumento das submissões, pressão sobre a revisão por pares) e registra que textos produzidos por LLM podem ser cientificamente persuasivos a ponto de enfraquecer os sinais tradicionais de qualidade textual. O segundo capítulo, “Ciência Aberta e editoração científica”, retoma a definição da Recomendação da Unesco sobre Ciência Aberta e as iniciativas da década de 1990, como a Open Archives Initiative, para sustentar que o editor deixa de ser apenas gestor de fluxos e passa a exercer “função estratégica de governança do conhecimento” (p. 13).

O terceiro capítulo é o núcleo analítico. Os autores identificam três pontos de entrada da IA: a triagem e as rotinas de integridade (similaridade textual, forense de imagens, checagens estatísticas como statcheck e o teste GRIM); o apoio à avaliação, em que modelos de linguagem resumem manuscritos, sintetizam literatura com RAG ou, em arquiteturas multiagente como o MARG, geram pareceres consolidados; e a normalização e a qualificação de metadados. A observação mais fina está no segundo ponto: quando o sistema passa a sintetizar e comentar, deixa de ser instrumento de checagem e passa a atuar “como mediador de interpretação” (p. 18). O Quadro 1 (pp. 19-20) sistematiza essa cartografia, relacionando, para cada etapa, o uso de IA, o que ela automatiza, o risco principal, a medida de governança sugerida e as fontes. Dele decorrem três tarefas: delimitar o que conta como uso substantivo, definir critérios de validação para cada sinal automatizado e registrar evidências e versões para auditoria (p. 21). O capítulo trata

ainda da injeção de instruções ocultas em manuscritos, com evidências experimentais de que texto invisível pode elevar recomendações automatizadas de aceitação de 0% para quase 100% (p. 21), e encerra com duas seções: uma sobre fundamentos éticos e marcos normativos (Unesco, Andifes) e outra sobre políticas autorais e metadados. Nesta, apoiados na tipologia de Arakaki e Arakaki, os autores mostram como metadados de autenticação, proveniência, descrição e direitos sustentam atribuição e auditoria, e afirmam que o ponto prático "não é permitir ou proibir ferramentas, mas explicitar responsabilidades e limites" (p. 25).

O quarto capítulo define governança editorial como conjunto de marcos éticos e políticas explícitas que orientam decisões, e articula o cenário brasileiro (relatório da CAPES, Guia de Boas Práticas Científicas da USP, Plano Brasileiro de Inteligência Artificial, resoluções da UNIFESP e da UFOP) ao internacional (Unesco, OCDE, AI Act europeu, COPE, ICMJE, WAME e políticas de editoras). Discute também os limites de instrumentos técnicos de triagem, como *watermarking* e *honeypots*, e propõe que o letramento em IA, ao explicitar usos aceitáveis e formas de declaração, "passa a operar como mecanismo de governança" (p. 35).

Os capítulos 5 e 6 dão ao livro sua feição mais singular. O quinto traduz princípios em quadros organizados por etapa (submissão, avaliação, edição de texto, curadoria de metadados e publicação, pós-publicação), cada item com condição de acionamento, ação, evidência mínima, risco endereçado e observação prática. Os autores advertem que o formato não cria obrigações nem substitui normas institucionais, e fixam um denominador comum: "decisões editoriais precisam permanecer humanas, justificadas e rastreáveis" (p. 37). Merecem destaque a "higienização do pacote para avaliação", que inclui a remoção de instruções ocultas e texto invisível; a proibição de terceirizar o juízo editorial a sistemas automatizados; o *checklist* de publicação do tipo *go/no-go*; a regra de que "o que aparece no artigo precisa existir e coincidir nos metadados" (p. 47); e a gradação dos instrumentos de atualização do registro, da mudança inclusiva de nome à retratação e à remoção, com vínculos bidirecionais e uso do Crossmark. O sexto capítulo converte esses quadros em protocolos passo a passo, em dois níveis: um mínimo, viável para equipes com poucos recursos, e um avançado, dependente de apoio técnico (p. 53). Há orientações concretas para registrar evidências no próprio

OJS, inspecionar camadas ocultas, aplicar sondas em pareceres e identificar referências alucinadas, cujos padrões comuns incluem “autor genérico, periódico inexistente, ano incoerente, título plausível demais” (p. 58).

As considerações finais reafirmam que a proposta não é “um modelo rígido ou universal”, mas um referencial adaptável (p. 67), e sintetizam o argumento central: práticas responsáveis dependem “menos do grau de sofisticação da tecnologia e mais da capacidade institucional de orientar, acompanhar e revisar seus próprios processos” (p. 67).

Avaliação crítica

O principal mérito do livro é a passagem do princípio à rotina. A literatura sobre IA e publicação científica tem produzido declarações de princípios que concordam no essencial (responsabilidade humana, vedação da autoria por IA, transparência), mas que pouco orientam o editor diante de um parecer suspeito ou de um pedido de mudança de autoria após o aceite. Ao estruturar cada recomendação com condição de acionamento e evidência mínima, os autores aproximam a governança editorial de uma lógica de controle auditável, sem cair no automatismo. A escolha de ancorar os procedimentos no OJS, plataforma predominante nos periódicos ibero-americanos, aumenta a utilidade imediata da obra para o público desta revista.

Um segundo mérito, de natureza epistemológica, é o tratamento das saídas automatizadas como sinais a interpretar e não como veredictos. O percentual de similaridade, o alerta de duplicação de imagem, a marca d’água detectada num parecer: todos aparecem como indícios que exigem contextualização e registro do critério humano aplicado. A posição é coerente com a tradição da Ciência da Informação de distinguir dado, evidência e julgamento, e protege os periódicos tanto da confiança ingênua nas ferramentas quanto de acusações indevidas.

O terceiro mérito, mais próximo das minhas pesquisas, é o lugar dado aos metadados. O livro os trata como mecanismo de transparência, rastreabilidade e responsabilização, capaz de registrar origem, intervenções humanas e automatizadas, versões e fluxos decisórios (p. 27). É exatamente aí que vejo o primeiro ponto a desenvolver. As declarações de uso de IA são previstas como

campo estruturado no sistema e como texto no manuscrito, em Métodos, Agradecimentos ou seção de transparência (p. 38). Como texto livre, porém, permanecem opacas para agregadores, indexadores e serviços de descoberta. Se a regra é que o que aparece no artigo deve coincidir nos metadados, o passo seguinte é que essas declarações sejam também legíveis por máquina: modeladas com vocabulários abertos, como a ontologia de proveniência PROV-O do W3C (Lebo, Sahoo, e McGuinness, 2013), que distingue entidades, atividades e agentes, inclusive agentes de software; articuladas aos papéis da taxonomia CRediT; e exportadas em formatos como o JATS (National Information Standards Organization, 2021) e os depósitos de DOI. Isso realizaria, na escala da web de dados (Berners-Lee, Hendler, e Lassila, 2001), o princípio de acionabilidade por máquina que os princípios FAIR (Wilkinson et al., 2016) já consagraram para dados de pesquisa, e permitiria perguntar a uma base, por exemplo, quais artigos declararam uso de IA generativa na análise de dados ou na tradução.

O segundo ponto decorre do anterior e diz respeito à direção do fluxo informacional. O livro trata a RAG sobretudo como ferramenta empregada por avaliadores e editores, com o risco bem identificado de produzir "sensação de verificabilidade sem checagem" (p. 20). Há, porém, um sentido inverso, que me parece decisivo: a literatura publicada é o corpus que LLMs e sistemas RAG consomem. Quando a geração é condicionada à recuperação de documentos (Lewis et al., 2020), a confiabilidade da resposta passa a depender da qualidade, da organização e da curadoria desse corpus. A governança de pós-publicação proposta pelos autores (retratações sinalizadas, expressões de preocupação, metadados relacionais do tipo "corrige/é corrigido por", Crossmark) torna-se, assim, condição a montante da confiabilidade da IA: um artigo retratado cujo status não é legível por máquina continua disponível para alimentar respostas. Explicitar essa dupla face, a de que governar a IA na editoração é também governar o conhecimento que a IA consumirá, ampliaria o alcance do livro e reforçaria o papel dos periódicos e dos profissionais da informação como infraestrutura crítica para uma IA confiável.

O terceiro ponto refere-se aos agentes de IA. A obra é atenta ao estado da arte ao incluir sistemas multiagente e a injeção de instruções ocultas, mas a difusão de agentes que leem arquivos, navegam, usam ferramentas e executam tarefas com autonomia amplia a superfície de risco. A injeção indireta de instruções foi descrita

justamente como vulnerabilidade de aplicações que processam conteúdo não confiável (Greshake et al., 2023), e um manuscrito submetido é, por definição, conteúdo não confiável. A defesa proposta no nível mínimo, de natureza manual (selecionar todo o texto, clicar em caixas e figuras, percorrer áreas vazias, p. 55), é razoável como ponto de partida, mas não escala. Ironicamente, parte dessa governança poderia ser automatizada com ferramentas determinísticas e transparentes, sem IA generativa: verificar cor, tamanho e posição de fonte em arquivos DOCX ou PDF é tarefa trivial para um *script* e poderia ser oferecida como *plugin* aberto para o OJS. Sobre os *honeypots*, os próprios autores reconhecem a ambivalência ética, já que a técnica defensiva é a mesma usada para manipular avaliações. Acrescento duas ressalvas: a instrução oculta em microfonte ou cor de fundo pode ser lida por leitores de tela, o que tensiona o compromisso com a acessibilidade que o livro defende em outras etapas (pp. 46, 49); e a sonda tende a detectar apenas usos ingênuos, pois agentes que pré-processam ou resumem o arquivo podem descartá-la, produzindo falsa sensação de controle. Se adotada, convém que sua existência seja informada na política para avaliadores.

O quarto ponto diz respeito à confidencialidade. A vedação ao envio de manuscritos a ferramentas “sem salvaguardas”, incluindo serviços comerciais que retêm dados ou os usam para treinamento (p. 41), é correta, mas a obra deixa em aberto o que conta como salvaguarda. Modelos de pesos abertos executados em infraestrutura institucional local constituem um caminho intermediário que preserva o sigilo, permite registrar versões e configurações e dialoga com a soberania digital do Plano Brasileiro de Inteligência Artificial, que o próprio livro menciona (pp. 29-30). Critérios mínimos para essa avaliação (onde o processamento ocorre, quais dados são retidos, que registros são gerados) seriam um complemento valioso aos quadros do capítulo 5.

Há, ainda, limites de método. Os quadros resultam de elaboração própria a partir da literatura e da experiência dos autores, sem que se explicite o procedimento de construção ou se apresente validação junto a editores. O prefácio reconhece que a obra representa “o estado atual” e inicia uma discussão (p. 7), e a divisão em níveis mínimo e avançado responde em parte à realidade dos periódicos pequenos, frequentes na América Latina e muitas vezes conduzidos por um único editor com apoio voluntário. Mesmo o nível mínimo, contudo, prevê registros em quase todas

as etapas, e seu custo operacional precisa ser medido. Estudos empíricos em periódicos de diferentes portes e áreas, por exemplo nas redes SciELO, Redalyc e Latindex, permitiriam calibrar o equilíbrio entre rigor e viabilidade, além de acompanhar a rápida obsolescência das ferramentas citadas.

Por fim, penso que esta é uma obra que permite já pensar na próxima edição, visto que os modelos baseados em LLM e inteligência artificial mudam com o passar do tempo.

Considerações finais

Inteligência artificial na editoração científica em Ciência Aberta oferece aos editores ibero-americanos algo que ainda faltava em língua portuguesa: um referencial que articula marcos éticos e normativos a procedimentos executáveis, situado na realidade institucional dos periódicos da região e na infraestrutura que eles efetivamente usam. Sua leitura é recomendável para editores, equipes editoriais, avaliadores, gestores de portais de periódicos e pesquisadores da comunicação científica.

Para a Ciência da Informação, o livro confirma que os desafios trazidos pela IA à publicação científica são, em grande medida, problemas informacionais: de proveniência, de representação, de curadoria e de atualização do registro. A agenda que ele abre, e que procurei indicar, inclui transformar declarações e status editoriais em metadados interoperáveis, reconhecer os periódicos como produtores dos corpus de que dependem os sistemas generativos e preparar o fluxo editorial para agentes autônomos. Ao deslocar o foco da sofisticação tecnológica para a capacidade institucional de governar processos, os autores oferecem uma base sólida para esse trabalho.

Referências bibliográficas

Berners-Lee, T., Hendler, J., e Lassila, O. (2001). The semantic web. *Scientific American, 284*(5), 34-43.

Greshake, K., Abdelnabi, S., Mishra, S., Endres, C., Holz, T., e Fritz, M. (2023). Not what you've signed up for: compromising real-world LLM-integrated applications with indirect prompt injection. In *Proceedings of the 16th ACM Workshop on Artificial Intelligence and Security* (pp. 79-90). New York, NY: ACM. https://doi.org/10.1145/3605764.3623985

Lebo, T., Sahoo, S., e McGuinness, D. (Eds.). (2013). *PROV-O: the PROV ontology*. W3C Recommendation. Recuperado de https://www.w3.org/TR/prov-o/

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N.,… Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. In *Advances in Neural Information Processing Systems 33* (pp. 9459-9474). Red Hook, NY: Curran Associates.

National Information Standards Organization. (2021). *JATS: Journal Article Tag Suite, version 1.3* (ANSI/NISO Z39.96-2021). Baltimore, MD: NISO.

Vechi, B. D., Shintaku, M., Maia, M. A. Q., e Sousa, R. P. M. de. (2026). *Inteligência artificial na editoração científica em Ciência Aberta: riscos, integridade científica e governança editorial*. Brasília, DF: Editora Ibict. https://doi.org/10.22477/9788570132253

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data, 3,* 160018. https://doi.org/10.1038/sdata.2016.18