
Desvendando a ‘visão de Deus’ Reflexões éticas e epistemológicas no uso de IAGs para Pesquisa em Comunicação¹

Luiza B Bucker²
ESPM, São Paulo, SP

RESUMO

A evolução das tecnologias comunicacionais e o avanço das Inteligências Artificiais Generativas (IAGs) têm gerado debates significativos sobre seu papel na pesquisa acadêmica. Este artigo propõe uma reflexão teórico-metodológica, situando a discussão no campo da filosofia da informação conforme Floridi (2013, 2014). Realizamos uma análise crítica de literatura para identificar definições, vieses, limitações e problemas éticos das IAGs, destacando a responsabilidade do pesquisador-autor. Concluímos que, para garantir a qualidade e a ética na pesquisa, é essencial que os pesquisadores reconheçam as limitações das IAGs e guiem seu uso de forma consciente e crítica.

PALAVRAS-CHAVE: Inteligência Artificial Generativa; epistemologia; ética na pesquisa; comunicação; filosofia da informação.

CORPO DO TEXTO

INTRODUÇÃO

A evolução das tecnologias comunicacionais e o avanço das Inteligências Artificiais Generativas (IAGs) têm gerado debates significativos sobre o papel dessas ferramentas na pesquisa acadêmica. Enquanto Russel e Norvig (2010) defendem a IA como um agente racional, capaz de fazer ‘a coisa certa’ a partir do que sabe, Crawford (2021) critica essa visão, apontando para as implicações políticas e sociais inerentes ao uso da IA.

Defenderemos, no primeiro momento, a sugestão da criação de uma filosofia da informação (FLORIDI, 2013), campo em que enquadraremos essa pesquisa voltada à epistemologia da comunicação em tempos de inteligência artificial. Fizemos uma varredura em principais artigos publicados na área e elencamos as principais definições, vieses, limitações, problemas éticos e sugestões para o uso de tais ferramentas na pesquisa acadêmica.

¹ Trabalho apresentado no GP Teorias da Comunicação, XIX Encontro dos Grupos de Pesquisas em Comunicação, evento componente do 47º Congresso Brasileiro de Ciências da Comunicação.

² Mestranda em Comunicação e Práticas do Consumo, ESPM-SP, email: luizabucker@gmail.com.

Russel e Norvig (2010) apontam para a diferenciação entre IA fraca e forte, máquinas podem agir como se fossem inteligentes e que as máquinas estão realmente pensando, respectivamente. Para tais teorias, ainda não há consenso, entretanto, consideram o fato de uma IA fraca ser possível. Esse entendimento corrobora para a linha de pensamento sobre uma racionalidade da IA, que é contraposta por outro entendimento.

Crawford (2021) defende que a inteligência artificial não é nem inteligente, nem artificial, no sentido de que dependem de um conjunto de estruturas políticas e sociais, sendo programadas para servir interesses dominantes, fugindo de um domínio puramente técnico. Aqui cabe ressaltar que um dos pesquisadores citados anteriormente estava como diretor de pesquisa do Google à época que escreveu o livro. Suas colocações técnicas não perdem validade, mas devemos contrapô-las com visões mais gerais do assunto. A autora coloca,

I use AI to talk about the massive industrial formation that includes politics, labor, culture, and capital. When I refer to machine learning, I'm speaking of a range of technical approaches (which are, in fact, social and infrastructural as well, although rarely spoken about as such). But there are significant reasons why the field has been focused so much on the technical — algorithmic breakthroughs, incremental product improvements, and greater convenience. The structures of power at the intersection of technology, capi-tal, and governance are well served by this narrow, abstracted analysis. To understand how AI is fundamentally political, we need to go beyond neural nets and statistical pattern recognition to instead ask what is being optimized, and for whom, and who gets to decide. Then we can trace the implications of those choices (Crawford, 2021, p.8,9).

Essa visão é construída, segundo ela, de forma a criar uma narrativa de que a inteligência artificial seria capaz de observar movimentos humanos, comunicacionais e de trabalho, o que ela denomina de 'God's-eye'. Uma visada colonialista que concentra o poder do campo da inteligência artificial, determinando como o mundo é medido e definido, enquanto nega sua atividade política (Crawford, 2021).

Diversos estudos denunciam os vieses produzidos tanto pela base de dados quanto para a forma como os algoritmos são programados. Apresentadas as iniciativas para melhoria da qualidade dos dados, fruto de reflexões sociais sobre seu uso, colocamos que, em campo macro, sugere-se a composição de uma filosofia da informação, cujo objeto é a natureza e os princípios básicos da informação, além de suas consequências éticas (FLORIDI, 2013). Tal filosofia considera que a sociedade da informação surgiu desde a invenção da escrita, pintura e mídia de massa, mas aponta a transição para um mundo

onde a capacidade das ICTs ultrapassou as capacidades de gravar e transmitir informação, evoluindo para capacidades de processamento. Por isso, na condição hiper histórica atual, se faz necessário a concepção de uma filosofia capaz de suprir o déficit conceitual causado pelas ICTs (FLORIDI, 2014).

Esse artigo é uma empreitada no sentido de refletir sobre autoria na pesquisa e criar novos vocabulários que podem servir de ferramenta, ou apenas indicativos de necessidade, para esse novo momento social com profundo impacto no campo da comunicação.

METODOLOGIA

Realizamos uma análise crítica de literatura, revisando artigos e livros que abordam a ética, as limitações e os vieses das IAGs. Utilizamos a abordagem de Floridi (2013, 2014) para situar nossa discussão no campo da filosofia da informação, investigando as implicações epistemológicas das IAGs na comunicação.

Inteligência artificial generativa é uma classe dentro dos modelos de IA que realiza mais do que apenas tarefas humanizadas, pode criar dados a partir do aprendizado de padrões e estruturas de outros dados. Funciona a partir de redes neurais e deep learning e pode gerar texto, imagens, músicas, entre outros (RAY, 2023).

Nesse cenário, a OpenAI lança o ChatGPT. Uma inteligência artificial generativa com processamento de linguagem natural (NLP), com o objetivo de criar uma versão sofisticada e versátil de um Generative Pre-trained transformer (modelos que leem e produzem textos com dados disponíveis online), capaz de realizar diversas tarefas (RAY, 2023). A empresa teve êxito, em estudo recente, o ChatGPT foi capaz de passar em provas da faculdade de direito de Minnessota, ainda que com notas ruins (CHOI et al, 2023).

O aprendizado profundo característico das IAGs cria uma ‘caixa preta’ na medida em que várias camadas de redes neurais se sobrepõem umas às outras, tornando as decisões tomadas pela ferramenta um mistério. Na busca pelo mapeamento dessas respostas, surge um movimento para criação de inteligências artificiais explicáveis (Explainable AI - XAI) (ALVES, 2022).

quando um algoritmo produz uma decisão incorreta – ou mesmo correta, mas apoiada em premissas falsas –, estaremos diante da falibilidade, situação em que o sistema de IA não opera da maneira desejável, seja por

razões ligadas ao design do algoritmo, seja pela forma como os dados são codificados, coletados, selecionados ou usados para treinar o algoritmo. Muitas vezes, a falibilidade tem efeitos inócuos. No entanto, quando uma falha produzida por algoritmos de IA afeta grupos ou indivíduos, potencialmente gerando resultados tendenciosos ou discriminatórios, ela adquire uma dimensão social, razão pela qual passa a ser tratada como viés algorítmico (Alves, 2022, p. 355).

Ray (2023) também discute as várias preocupações que emergem do uso do ChatGPT, destacando uma gama de vieses e limitações inerentes. Listamos todos os vieses e algumas limitações para que fique claro que a visão de que a IA tem os ‘olhos de Deus’ está equivocada. O autor ainda aponta para os questionamentos que levantamos no artigo:

Ethical Use of AI-generated Content: The widespread use of AI-generated content raises ethical questions about the appropriate contexts and applications for such content. For example, using AI-generated content in journalism or academic research may raise concerns about authenticity, integrity, and the potential for plagiarism. Establishing ethical guidelines and best practices for the use of AI-generated content in various contexts can help mitigate these concerns and ensure responsible use (RAY, 2023, p.146).

Buscamos mapear o processo de pesquisa acadêmica para compreender os riscos da utilização das ferramentas na jornada científica.

FUNDAMENTAÇÃO TEÓRICA

Para a análise de discurso de linha francesa, autor é aquele que exerce uma função discursiva-enunciativa, moldado pelo já-la, ou seja, pelo interdiscurso ou memória social que afeta o intradiscurso. Não se pode naturalizar o que foi dito como a única maneira de se dizer, ressaltando a relevância daquilo que deixou de se dizer, o que deslegitima a neutralidade do discurso e coloca o autor como aquele que escolhe o dizer e o não dizer (ORLANDI, 2015). Autor é aquele que organiza argumentos (PACÍFICO, 2011), gerencia a expectativa dos leitores (DUNLEAVY, 2003) e dá a ilusão de começo, meio e fim para um texto (BARRICHELLO, 2016).

A partir dessa definição de autoria, caso a inteligência artificial generativa funcionasse como uma fonte segura de informações, ela poderia exercer o papel de autor. Porém, durante todo o artigo demonstramos as distorções geradas pela ferramenta. Somadas à essas reflexões, trazemos o colonialismo de dados na computação afetiva (CORTIZ e FERREIRA, 2022). Ainda que a IAG seja utilizada como um autor, é um

autor treinado por uma visão distorcida do mundo e treinada para convencer humanos sobre a assertividade de seu resultado.

Dado que existe essa falsa narrativa de que a inteligência artificial fornece dados corretos e de que a filosofia da informação ou iniciativas epistemológicas para desvendar tal tecnologia ainda se encontram insipientes, é necessário que sejam voltados esforços para compreender os riscos e limites do uso de tais tecnologias. Algo ainda mais importante do ponto de vista da produção acadêmica atual, que passa pelo processo entendido como ‘produtivismo acadêmico’, relacionado aos incentivos governamentais que medem excelência por critérios relacionados a quantidade produzida, apresentando impacto na qualidade das pesquisas (BIANCHETTI; VALLE, 2014). Essa busca por alta produtividade em tempos de uma tecnologia que pode otimizar os esforços de pesquisa pode gerar vieses, que caso desapercibidos, alteram o rumo científico.

Ainda assim, colocadas as limitações, cabe pontuar como a IAG pode auxiliar na otimização do tempo do pesquisador. Para isso, é necessário enfatizar seu papel como ‘pesquisador-autor’.

Utilizamos as premissas de Barrichello (2016) para nos guiar rumo à compreensão desse conceito. A autora à época, já dispõe do fato de que as plataformas digitais facilitam a possibilidade de acesso aos textos sobre os temas da pesquisa e sua implicação em problemas como plágio ou citação de vários autores sem contextualizá-los, assumindo a voz do autor, que não organiza o argumento da tese. Se é verdade que a plataforma de IAG são um facilitador para a revisão de literatura, também podemos assumir que esses problemas foram agravados.

Delimitamos os processos de ‘pesquisa da pesquisa’, elaborando o estado da arte e observando as lacunas para retomar a questão de pesquisa e delimitar as palavras chave, apontando que o uso de IAGs nunca é neutra e irá privilegiar assuntos, cabendo ao pesquisador ampliar sua fonte de pesquisa. Pode-se utilizar para organizar citações (RAMOS, 2023), mas cabe ao pesquisador formular argumento próprio e revisar tais adiantamentos, distinguindo o trabalho grosso das IAGs com o pente fino do autor-pesquisador. Ressaltamos a responsabilidade do pesquisador-autor em elaborar a articulação do diálogo das questões com os autores, garantindo a cientificidade, pois sempre que a linha teórica-metodológica escolhida é tensionada com o objeto, o autor de posiciona no trabalho (BARRICHELO, 2016).

ANÁLISE E RESULTADOS

Identificamos que as IAGs, apesar de suas capacidades avançadas, estão sujeitas a diversos vieses, como o viés cultural, linguístico, de gênero e racial (Ray, 2023). Esses vieses afetam a produção acadêmica, especialmente na revisão de literatura e na formulação de argumentos. Ressaltamos a importância de um uso crítico e consciente dessas ferramentas, com ênfase na revisão por pares e na transparência sobre as etapas da pesquisa realizadas com auxílio de IAGs.

O termo ‘meta-autor’ foi cunhado para designar um autor que ‘estipula o tema da pesquisa e outros dados acerca do escopo e extensão da investigação, e delega em seguida ao algoritmo a tarefa de vasculhar bancos de dados e de analisar e organizar as informações coletadas’ (ARAÚJO, 2016, p.95). Nesse caso, a pesquisa teria um duplo viés (bias), o do pesquisador, que manipula seu discurso conforme sua bagagem, inserção cultural e interesses e o da ferramenta, que faz o mesmo a partir de sua base de dados, programação e interesses. É, portanto, imprescindível deixar claro qual etapa do trabalho foi realizado por IAs para que os riscos iminentes de sua utilização pontuados aqui possam ser mitigados a partir da revisão por pares. Na próxima sessão, elaboramos as etapas de pesquisa, apontando o papel da autoria em cada etapa do trabalho. Assim, busca-se criar uma espécie de ‘guia introdutório’ para que seja verificada a partir da indicação de que, por exemplo, o texto foi redigido por humanos, revisado por Inteligência Artificial, ou a revisão bibliográfica foi feita inteiramente ou parcialmente (descrever para quais processos foi utilizado) e assim para toda a pesquisa.

Ainda que a pesquisa seja revisada por pares ao indicar corretamente em quais etapas do processo foram utilizadas e o autor retome sua autoria, as inteligências artificiais treinadas com base de dados disponíveis online não necessariamente partem de informações conscientemente disponibilizadas pelo seu criador para tais fins. Então, fica o questionamento: até que ponto a utilização de argumentos produzidos por IAGs pode resultar em um certo ‘parasitismo do conhecimento’, em que o beneficiário prejudica o autor fonte ao recuperar sua autoria? Essa concepção de parasitismo do conhecimento se distingue do plágio convencional, na medida em que não se trata de uma cópia direta, deliberada e não autorizada, mas da maneira como a IA cria a partir de padrões e estruturas existentes, sem sua devida fundamentação; criando conteúdo novo, mas

fundamentado em informações não verificadas e apropriadas de forma automatizada e potencialmente não éticas. Enquanto o plágio é responsabilidade do autor, o 'parasitismo do conhecimento' aciona responsabilidade tanto do usuário da IA quanto dos criadores das IAGs, uma vez que a ferramenta pode gerar conteúdo sem a transparência necessária sobre as fontes e os processos envolvidos.

CONCLUSÃO

Concluimos que, para garantir a qualidade e a ética na pesquisa, é fundamental que os pesquisadores assumam um papel ativo na utilização de IAGs, reconhecendo suas limitações e vieses. Para que a autoria seja garantida, o pesquisador deve ter a capacidade de guiar a ferramenta de pesquisa, qualificando suas respostas. Caso mantenha tal postura e siga os critérios de utilização das IAGs como ferramenta de pesquisa, ciente de suas limitações e colaborando com necessária revisão de possíveis vieses, ele terá um caminho mais proveitoso em sua pesquisa, assumindo os riscos da nova ferramenta de pesquisa. Buscou-se assim, contribuir para o campo da filosofia da informação, conforme proposta por Floridi, que oferece um framework necessário para entender e mitigar os impactos das IAGs na comunicação.

REFERÊNCIAS

- ARAUJO, Marcelo. "O uso de inteligência artificial para a geração automatizada de textos acadêmicos: plágio ou meta-autoria?" *Logeion: Filosofia da Informação* 3.1 (2016): 89-107.
- BAIDOO-ANU, D.; OWUSU ANSAH, L. "Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning". (2023) Available at SSRN 4337484.
- BARICHELLO, Eugenia Maria Mariano Da Rocha et al. "Reflexões Sobre A Formação Do Pesquisador - A Autoria Na Elaboração De Uma Tese." In: MOURA, Cláudia Peixoto De; LOPES, Maria Immacolata Vassalo de, et al. (Org.). *Pesquisa Em Comunicação: Metodologias E Práticas Acadêmicas*. Porto Alegre: Edipucrs, 2016.
- BENDER, E.M.; GEBRU, T.; MCMILLAN-MAJOR, A.; SHMITCHELL, S. "On the dangers of stochastic parrots: Can language models be too big?" In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency, Virtual*, 3–10 March 2021; pp. 610–623.

BIANCHETTI, Lucídio; VALLE, Ione Ribeiro. "Produtivismo acadêmico e decorrências às condições de vida/trabalho de pesquisadores brasileiros e europeus.". 2014.
<https://doi.org/10.1590/S0104-40362014000100005>

BROWN, Tom et al. "Language models are few-shot learners." *Advances in neural information processing systems* 33 (2020): 1877-1901.

BUOLAMWINI, Joy; COSTANZA-CHOCK, Sasha; RAJI, Inioluwa Deborah. "Who Audits the Auditors? Recommendations from a field scan of the algorithmic auditing ecosystem." In *2022 ACM Conference on Fairness, Accountability, and Transparency*. 1571–1583.

BUOLAMWINI, Joy. "Limited Vision: The Undersampled Majority." 2017.

CHOI, J.H.; HICKMAN, K.E.; MONAHAN, A.; SCHWARCZ, D. "ChatGPT Goes to Law School." SSRN 2023, 4335905. Available online: <https://ssrn.com/abstract=4335905>

CORTIZ, D.; FERREIRA DOS SANTOS, J.C. "Computação Afetiva: entre as limitações técnicas e os desafios do colonialismo de dados." *Revista Fronteiras*. 2022 Sep 1;24(3).

CRAWFORD, Kate. "The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence." 2021.

FLORIDI, Luciano. "The Ethics of Information". 2013.

https://books.google.com.br/books?hl=pt-BR&lr=&id=_XHcAAAAQBAJ&oi=fnd&pg=PP1&dq=Luciano+Floridi&ots=f_fL7_VvUZ&sig=jSxIf52kk-4g8PWbvyy4Ysz2xHU#v=onepage&q=Luciano%20Floridi&f=false

FLORIDI, Luciano. "The Fourth Revolution: How the Infosphere is Reshaping Human Reality".
https://books.google.com.br/books?hl=pt-BR&lr=&id=65eAAwAAQBAJ&oi=fnd&pg=PP1&dq=Luciano+Floridi&ots=n_87zhPyCK&sig=c0nyMqNRTjBzbaqQxb3GVrfRmbw#v=onepage&q=Luciano%20Floridi&f=false

FERREIRA, William. "Distorted Lenses: Unveiling Gender Stereotypes in Images Generated by Artificial Intelligence"; 2024.

ORLANDI, Eni. "Análise de discurso, princípios e procedimentos.". Pontes; 5ª edição, 2015.

Para começar um projeto de pesquisa. (2005). "Comunicação & Educação", 10(3), 288-296.
<https://doi.org/10.11606/issn.2316-9125.v10i3p288-296>

QUALITY ASSURANCE AGENCY FOR HIGHER EDUCATION – QAA. "The framework for higher education qualifications in England, Wales, and Northern Ireland.". Gloucester, 2001.

RAMOS, Anália Saraiva Martins. "Inteligência Artificial Generativa baseada em grandes modelos de linguagem - ferramentas de uso na pesquisa acadêmica." [S. l.], maio 2023. DOI: 10.1590/SciELOPreprints.6105. Disponível em:
<https://preprints.scielo.org/index.php/scielo/preprint/view/6105/version/6463> Acesso em: 16 de março de 2024.

RAY, P.P. "ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations, and future scope.". *Internet Things Cyber-Phys. Syst.* 2023, 3, 121–154.

ROZADO, D. "The political biases of chatgpt.". *Soc. Sci.* 2023, 12, 148.

RUSSELL, Stuart J.; NORVIG, Peter; DAVIS, Ernest. "Artificial Intelligence: A Modern Approach.". Upper Saddle River, NJ, Prentice Hall, 2010.

World Economic Forum Global Future Council on Human Rights 2016–2018. "How to Prevent Discriminatory Outcomes in Machine Learning"; 2018.

https://www3.weforum.org/docs/WEF_40065_White_Paper_How_to_Prevent_Discriminatory_Outcomes_in_Machine_Learning.pdf

ZHANG, Zhuosheng; ZHANG, Aston; LI, Mu; SMOLA, Alex. "Automatic chain of thought prompting in large language models.". arXiv preprint arXiv:2210.03493 (2022).