

Ética e Inteligência Artificial: explorando indicadores da produção científica

Ednéia Silva Santos Rocha

Universidade de São Paulo, Programa de Pós-Graduação em Ciência da Informação, São Paulo, SP, Brasil

ORCID: <https://orcid.org/0000-0003-1478-6828>
edneia@usp.br

Diogo Roberto da Silva Andrade

Universidade do Estado de Santa Catarina, Programa de Pós-Graduação em Gestão da Informação, Florianópolis, SC, Brasil

ORCID: <https://orcid.org/0000-0002-8392-4481>
didts@hotmail.com

Jean Carlos Ferreira dos Santos

Universidade de São Paulo, Núcleo de Informação e Coordenação do Ponto BR, São Paulo, SP, Brasil

ORCID: <https://orcid.org/0000-0003-3860-9746>
jeanferreira@usp.br

DOI: <https://doi.org/10.26512/rici.v19.n1.2026.57464>

ARTIGOS DE REVISÃO

Recebido/Recibido/Received: 2025-03-05

Aceito/Aceptado/Accepted: 2025-12-08

Publicado/Publicado/Published: 2026-02-20

Resumo

O artigo mapeia a produção científica sobre ética e inteligência artificial (IA) por meio de indicadores bibliométricos, analisando tendências e o estado atual do campo. Utilizando a base Scopus, devido à sua ampla cobertura e qualidade dos metadados, a pesquisa seguiu uma estratégia de busca específica para coletar e analisar os dados. Os resultados apontam um crescimento expressivo das publicações sobre ética e IA, com destaque para Estados Unidos, Reino Unido, Alemanha e Finlândia. No entanto, há baixa representatividade de países da América Latina, África e Ásia Central. A análise bibliométrica mostra que a maioria dos estudos está concentrada em periódicos e conferências renomadas, como a AAAI/ACM Conference on AI, Ethics, and Society e o *Journal of Business Ethics*. Os principais temas abordados incluem responsabilidade social dos desenvolvedores, transparência algorítmica, privacidade de dados e impactos sociais e econômicos da IA. Uma lacuna importante identificada é a escassez de discussões sobre o uso da IA na produção acadêmica, especialmente em relação a plágio, vies algorítmico e integridade científica. O estudo conclui que é essencial estabelecer diretrizes claras para o uso ético da IA garantindo transparência, responsabilidade e controle humano sobre a tecnologia.

Palavras-chave: Indicadores. Ética na informação. Inteligência artificial.

Ética e Inteligencia Artificial: explorando indicadores de producción científica

Resumen

El artículo mapea la producción científica sobre ética e inteligencia artificial (IA) a través de indicadores bibliométricos, analizando las tendencias y el estado actual del campo. Utilizando la base de datos Scopus, debido a su amplia cobertura y calidad de metadatos, la investigación siguió una estrategia de búsqueda específica para recolectar y analizar los datos. Los resultados indican un crecimiento significativo de las publicaciones sobre ética e IA, con énfasis en Estados Unidos, Reino Unido, Alemania y Finlandia. Sin embargo, hay baja representación de países de América Latina, África y Asia Central. El análisis bibliométrico muestra que la mayoría de los estudios se concentran en revistas y conferencias de renombre, como la Conferencia AAAI/ACM sobre IA, Ética y Sociedad y el *Journal of Business Ethics*. Los temas clave que se cubren incluyen la responsabilidad social de los desarrolladores, la transparencia algorítmica, la privacidad de los datos y los impactos sociales y económicos de la IA. Una brecha importante identificada es la falta de discusiones sobre el uso de IA en la producción académica, especialmente en relación con el plagio, el sesgo algorítmico y la integridad científica. El estudio concluye que es esencial establecer directrices claras para el uso ético de la IA, garantizando la transparencia, la responsabilidad y el control humano sobre la tecnología.

Palabras clave: Indicadores. Ética de la información, Inteligencia artificial.

Ethics and Artificial Intelligence: exploring scientific field indicators

Abstract

The article maps the scientific production on ethics and artificial intelligence (AI) through bibliometric indicators, analyzing trends and the current state of the field. Using the Scopus database, due to its broad coverage and quality of metadata, the research followed a specific search strategy to collect and analyze the data. The results indicate a significant growth in publications on ethics and AI, with emphasis on the United States, the United Kingdom, Germany, and Finland. However, there is low representation of countries in Latin America, Africa, and Central Asia. The bibliometric analysis shows that most studies are concentrated in renowned journals and conferences, such as the AAAI/ACM Conference on AI, Ethics, and Society and the Journal of Business Ethics. The main topics addressed include social responsibility of developers, algorithmic transparency, data privacy, and social and economic impacts of AI. An important gap identified is the lack of discussions on the use of AI in academic production, especially in relation to plagiarism, algorithmic bias, and scientific integrity. The study concludes that it is essential to establish clear guidelines for the ethical use of AI, ensuring transparency, accountability and human control over technology.

Keywords: Indicators. Information ethics. Artificial intelligence.

1 Introdução

A Inteligência Artificial (IA) tem sido tratada como um recurso estratégico para o desenvolvimento econômico, político e social desde meados do século XX. O ano de 1950 marca simbolicamente esse início com a proposição do “Teste de Turing”, elaborado por Alan Turing, que buscava responder se as máquinas poderiam pensar, inaugurando assim a base filosófica e computacional da IA moderna. Desde então, a evolução dos sistemas de IA tem se intensificado, alcançando inovações notáveis especialmente a partir da década de 2020, com a aplicação massiva em diversos setores informacionais e comunicacionais.

O conceito de Inteligência Artificial (AI, na língua inglesa) refere-se ao estudo de como se produzem máquinas que tenham algumas qualidades próximas às da mente humana, como

a habilidade de entender uma linguagem, reconhecer figuras, resolver problemas e aprender (Artificial Intelligence, 2023). A diferença entre a IA tradicional e a IA generativa está no uso de modelos mais avançados de aprendizado de máquina, como o *deep learning*. A IA generativa é capaz de simular processos de aprendizagem e tomada de decisão da mente humana e, assim, pode criar conteúdo original – como texto, imagens, vídeo, áudio ou código de software – em resposta a um prompt ou solicitação do usuário (Stryker; Scapicchio, 2024). Contudo, deve ser observada que tanto a autenticidade como a originalidade devem ser percebidas, refletidas e criticadas, pois atravessam questões sobre princípios éticos, da segurança da informação e da privacidade de dados e conteúdos gerados pelas IA generativas.

Nesse cenário, cresce a necessidade de refletir sobre as implicações éticas do uso da IA na vida cotidiana. A ética torna-se central no debate à medida que a IA deixa de ser apenas uma ferramenta técnica para se integrar às decisões humanas mais sensíveis, como saúde, justiça, educação e segurança. As produções acadêmicas e científicas podem sinalizar valores e conflitos políticos, econômicos, culturais e sociais que permeiam a construção de diretrizes éticas e regulatórias sobre o uso da IA.

O desenvolvimento da IA tem envolvido não apenas governos e pesquisadores, mas também empresas privadas, sobretudo as grandes corporações tecnológicas (*Big Techs*), que detêm poder econômico, computacional e político significativo. Tais empresas não apenas influenciam, mas muitas vezes moldam as políticas públicas e as regulamentações sobre IA em escala global (Zuboff, 2021; Crawford, 2021). Como destacam Benjamin (2019) e Pasquale (2015), essa centralização de poder pode perpetuar desigualdades estruturais e consolidar regimes de opacidade algorítmica. A ética da IA, portanto, não pode ser compreendida sem considerar os interesses econômicos e geopolíticos que atravessam sua formulação e implementação.

Considerando esse contexto, o objetivo geral deste estudo é produzir indicadores a partir da análise de bases de dados científicas, com foco na produção internacional sobre ética e IA. Para isso, definiram-se dois objetivos específicos: (1) mapear as publicações científicas sobre a relação entre ética e IA na base de dados Scopus; e (2) identificar tendências e padrões que revelem a evolução desse campo de pesquisa entre 2015 e 2022 — recorte temporal justificado pela emergência das IA generativas e pelo aumento significativo da produção acadêmica sobre o tema nesse período.

A importância da abordagem transdisciplinar, especialmente no campo da Biblioteconomia e Ciência da Informação (BCI), é ressaltada na medida em que este campo oferece recursos teóricos e metodológicos para compreender como a informação circula, é registrada e usada eticamente na sociedade contemporânea (Schneider, 2013). O estudo se

insere nessa interseção crítica entre informação, ética e tecnologia, buscando contribuir para o mapeamento das dinâmicas temáticas, epistemológicas e políticas que estruturam a produção científica sobre ética e IA.

Entende-se que a ética em IA envolve, entre outros aspectos sensíveis, discussões sobre transparência algorítmica, responsabilidade dos desenvolvedores, privacidade de dados, justiça social, vieses automatizados e as questões econômicas e políticas que perpassam o desenvolvimento tecnológico. Prates e Chiarini (2024) comentam que o cenário digital global é liderado por grandes corporações que controlam as plataformas digitais e, também, a tecnologias, incluindo os desenvolvimentos em torno da IA. Essas corporações, a maioria sediada nos EUA, detém grande poder nas dinâmicas econômicas e políticas. Compreende-se que o poder concentrando das corporações tecnológicas tem um papel significativo na disseminação e no uso generalizado de seus serviços e ferramentas nas diversas instâncias sócias. Nesse sentido, a análise sistemática da produção científica permite não apenas identificar tendências e lacunas do conhecimento, mas também fomentar o debate sobre a construção de diretrizes éticas e políticas institucionais que assegurem o desenvolvimento responsável da IA em benefício da sociedade de forma justa, equitativa e transparente. Cabe, ainda, problematizar o discurso de liderança tecnológica por parte de países em desenvolvimento, frequentemente presente em planos estratégicos nacionais (Cortiz, 2020), à luz das limitações estruturais desses contextos e da assimetria global de infraestrutura, dados e soberania tecnológica (Benjamin, 2019; Eubanks, 2018).

2 Ética em inteligência artificial: tópicos e questões emergentes

A década de 2020 marca uma corrida global para o avanço dos sistemas de Inteligência Artificial (IA), frequentemente descrita como uma tecnologia de alto valor estratégico. Países em diversas regiões – tanto desenvolvidos quanto em desenvolvimento – lançaram planos nacionais para posicionar-se como líderes na chamada economia digital (Cortiz, 2020). No entanto, autores como Benjamin (2019) e Zuboff (2021) alertam que essa narrativa de liderança tecnológica pode mascarar desigualdades profundas relacionadas à infraestrutura, à soberania sobre dados e ao acesso à inovação. Assim, embora o discurso político e técnico afirme o valor econômico e social da IA, sua apropriação é desigual e concentrada em poucos atores globais.

O saber científico e tecnológico, nesse contexto, é considerado um ativo geopolítico e cognitivo, perpassando dimensões técnicas, sociais e éticas que exigem abordagens críticas e interdisciplinares (Schneider, 2013). Arbix (2020) destaca que a capacidade da IA de resolver problemas complexos com precisão e baixo custo tem impulsionado sua disseminação, mas também elevado os riscos associados a usos não regulados, opacos e discriminatórios. Cortiz

(2020), por sua vez, chama atenção para as consequências inesperadas decorrentes da emergência dessas tecnologias, que desafiam os modelos regulatórios tradicionais.

Empresas privadas desempenham papel central nesse ecossistema. As chamadas *Big Techs* acumulam poder computacional, financeiro e político, influenciando diretamente a formulação de normas e diretrizes internacionais sobre ética em IA. Essa configuração concentra capacidades decisórias nas mãos de poucos atores, aprofundando o desequilíbrio entre países centrais e periféricos (Pasquale, 2015; Zuboff, 2021).

Frente a isso, várias organizações têm buscado estabelecer princípios normativos de uso ético da IA. O AI EthicsLab (2022), o AI NowInstitute (2023), a Partnership on AI (2023) e o Institute for Human-Centered Artificial Intelligence (HAI, 2021) são exemplos de iniciativas que propõem diretrizes sobre privacidade, transparência e responsabilidade. Na esfera política, a Comissão Europeia (2019) e a OCDE (2019) divulgaram recomendações para orientar governos e empresas, promovendo uma IA confiável e socialmente responsável.

A OCDE (2019), por exemplo, propõe conceitos fundamentais para a criação de políticas digitais, ou seja, um conjunto de normas que regulam os sistemas tecnológicos sob critérios de equidade, eficiência e direitos fundamentais. Diferenciam-se de políticas públicas, geralmente articuladas a programas estatais de prestação de serviços, e de políticas de Estado, associadas a compromissos de longo prazo que transcendem governos. No caso da IA, políticas digitais são frequentemente formuladas por coalizões híbridas de governos, empresas e universidades, o que levanta questionamentos sobre sua legitimidade democrática.

Entre os conceitos apresentados, destacam-se os de agência e supervisão humana: a agência refere-se à capacidade de ação deliberada dos sujeitos envolvidos nas decisões algorítmicas; já a supervisão humana diz respeito ao acompanhamento e avaliação crítica das decisões automatizadas, impedindo que a autonomia dos sistemas ultrapasse limites éticos e legais. Como destaca a própria OCDE (2019), a supervisão é condição fundamental para lidar com riscos de uso indevido intencional ou acidental de sistemas de IA.

Para melhor alinhamento das normativas em torno da ética e IA o comitê OCDE (2019) propôs conceitos para a criação de políticas digitais (Quadro 1), na pretensão de nortear e estimular a uniformidade nas questões que envolvam IA e ética.

Quadro 1 – Quadro conceitual

Termo	Conceito
Sistema de IA	Sistema baseado em máquina que, para objetivos explícitos ou implícitos, infere, a partir da entrada que recebe, como gerar saídas como previsões, conteúdo, recomendações ou decisões que podem influenciar ambientes físicos ou virtuais. Diferentes sistemas de IA variam em seus níveis de autonomia e adaptabilidade após

Termo	Conceito
	a implantação.
Ciclo de vida do sistema de IA	Normalmente envolve várias fases que incluem: planejar e projetar; coletar e processar dados; construir modelo(s) e/ou adaptar modelo(s) existente(s) a tarefas específicas; testar, avaliar, verificar e validar; disponibilizar para uso/implantar; operar e monitorar; e aposentar/descomissionar. Essas fases geralmente ocorrem de forma iterativa e não são necessariamente sequenciais. A decisão de aposentar um sistema de IA da operação pode ocorrer a qualquer momento durante a fase de operação e monitoramento.
Atores de IA	São aqueles que desempenham um papel ativo no ciclo de vida do sistema de IA incluindo organizações e indivíduos que implantam ou operam IA.
Conhecimento de IA	Alude às habilidades e recursos, como dados, código, algoritmos, modelos, pesquisa, conhecimento, programas de treinamento, governança, processos e melhores práticas necessárias para entender e participar do ciclo de vida do sistema de IA incluindo o gerenciamento de riscos.
Partes Interessadas (<i>stakeholders</i>)	Abrangem todas as organizações e indivíduos (partes interessadas) que sejam envolvidos ou afetados por sistemas de IA, direta ou indiretamente. Neste caso, os Atores de IA são um subconjunto de <i>stakeholders</i> .

Fonte: OCDE (2019, tradução nossa, grifo nosso).

Contudo, apesar da explosão de iniciativas interessadas em endereçar as questões e incertezas que emergem com a expansão dos sistemas de IA, preocupações com a conceituação e definição do que constitui a ética em IA e quais os requisitos éticos específicos, padrões técnicos e melhores práticas necessárias para a sua implementação tem emergido nos mais variados campos disciplinares acadêmicos-científicos e nos setores políticos, econômicos e sociais (Jobin; Ienca; Vayena, 2019).

Conforme menciona Cortiz (2020), os economistas têm se preocupado com a adoção em larga escala de IA, pois a longo prazo esta prática pode ter reflexos sociais e econômicos significativos, como mudanças na força de trabalho e desigualdades sociais. A Psicologia e Neurociência investigam as possíveis consequências que a IA pode trazer para a cognição e o comportamento (Cortiz, 2020). Na área jurídica, as preocupações compreendem o impacto significativo da IA no sistema judicial, na aquisição e violação de direitos (Cortiz, 2020). Os juristas também têm se debruçado sobre a necessidade de se estabelecer diretrizes éticas e regulamentações adequadas para garantir que a IA seja desenvolvida e usada respeitando os direitos e valores das pessoas e seus coletivos (Cortiz, 2020). Já nas Ciências Sociais, os estudiosos têm levantado preocupações com acentuação das desigualdades pelo processo de automação de decisões e serviços e seus possíveis efeitos excludentes e discriminatórios, bem como as reconfigurações das dinâmicas sociais e políticas decorrentes do largo uso de sistemas de inteligência artificial (Cortiz, 2020).

Ainda na esteira das pesquisas acadêmicos-científicos, Muller (2020) examinou os dilemas éticos que surgem com o avanço dos sistemas de IA e robótica, sobretudo aqueles decorrentes de certos usos dessas tecnologias. O autor também descreve as discussões em

curso sobre o desenvolvimento de uma ética para os próprios sistemas de IA, no contexto da "ética para as máquinas". Em tal contexto, considera-se a IA não apenas como meros objetos usados pelos seres humanos, mas sim como sujeitos éticos, e explora a ideia dos sistemas de IA como "agentes morais artificiais" que possuem direitos e responsabilidades. Menciona-se também as preocupações que envolvem a "singularidade" da IA, isto é, o desenvolvimento de uma possível futura superinteligência artificial capaz de superar a inteligência humana, cujos efeitos e controle são de difícil previsão e, portanto, preocupantes. Alguns acreditam que a superinteligência poderá trazer benefícios, enquanto outros apontam para inúmeros riscos associados a ela (Muller, 2020).

As discussões sobre os riscos de a IAM alcançar a superinteligência visam a necessidade de análises profundas sobre o comportamento dos sujeitos e dos coletivos que interpretam a IA como uma autoridade cognitiva sem falhas. Isto se dá, pois, o imaginário social compreende a computação como um sistema dotado de lógica e matemática que não permite falhas de execução de comandos para a geração de respostas que auxiliem nas tomadas de decisão. Neste sentido, estudos dos diversos campos frisam que não há neutralidade na IA, pois o seu *fazer* está alicerçado nos algoritmos projetados por humanos, bem como os *prompts* (Interpretador de comandos) de comando. Ainda assim, a própria ideia de que tecnologias seriam neutras – por operarem com base em lógica e dados – tem sido contestada. Winner (1980) argumenta que artefatos tecnológicos não são isentos de valores e podem incorporar intenções políticas ao moldar a organização da vida social. Latour (1999) amplia essa perspectiva ao demonstrar que objetos técnicos fazem parte de redes sociotécnicas e têm papel ativo na mediação das práticas humanas. Jasanoff (2004) reforça que o desenvolvimento científico e tecnológico está intimamente ligado a decisões políticas, jurídicas e culturais, sendo coproduzido com as instituições sociais.

Segundo Moraes (2019, p. 63): "na contemporaneidade, a hipótese de tais tecnologias já não se sustenta, pois elas incorporam em sua programação objetivos determinados pelos programadores segundo seus próprios interesses e propósitos". Continuando, um aspecto importante da ética em IA segundo Muller (2020) envolve pensar a IA como uma ferramenta desenvolvida e utilizada por humanos e considerar as questões éticas que emergem com os seus determinados usos. Isso tem incluído questões éticas¹ relativas à privacidade e manipulação, principalmente aquelas levantadas em razão das possibilidades de coleta e análise inteligente

¹ A ética pode ser descrita como princípios, valores, hábitos e costumes criados para se viver em sociedade, que se modificam pelo tempo, espaço e cultura (Cremonese, 2019). Por este motivo, o cenário atual das IA e redes sociais (*Instagram, Facebook, X/Twitter*) a economia da atenção deve também ser observada no que tange os sujeitos em relação com as informações no meio digital, ética e IA.

de dados em um contexto de economia da atenção e direcionamento do comportamento. Estas implicações éticas são crescentes por meio do uso de informações para manipular o comportamento *online* e *offline* dos usuários, minando a escolha racional autônoma, tanto no contexto de campanhas de *marketing* como de escolhas políticas e orientações ideológicas.

Diante dessa realidade emergente – induzida pelo capitalismo de vigilância e o neoliberalismo contemporâneo –, as liberdades civis e os direitos individuais se encontram progressivamente ameaçados pela interferência das empresas, vigilância estatal e menor proteção à privacidade em comparação ao período anterior à expansão das redes digitais (Zuboff, 2021). Observa-se o contexto dos buscadores *online* e plataformas de redes sociais, cujos algoritmos são desenvolvidos e controlados por *Big Techs* – empresas privadas como *Apple*, *Amazon*, *Microsoft*, *Meta (Facebook)* e *Alphabet (Google)* –, que detêm dados de grande parte da população que se conecta em rede, essas empresas utilizam seus recursos tecnológicos digitais para captar e vender dados, além disso ampliam as capacidades da IA. O paralelo virtual permite, dessa forma, que os sujeitos concebem o mundo, e as informações produzidas nele, por meio de controle algorítmico (Assange, 2015).

Garcia (2020) aponta que o emprego de técnicas de IA não só exacerba, mas perpetua distorções, como vieses raciais e desigualdades diversas. Para Danaher (2016), os sistemas de IA compõem uma estrutura de poder na qual estão sendo criados processos de tomada de decisão que restringem e limitam a participação humana. Ao mesmo tempo, fica cada vez mais difícil o usuário afetado entender como o sistema chegou a um determinado resultado, ou seja, o sistema mostra-se “opaco” para esse usuário – e possivelmente até mesmo para os seus desenvolvedores. Esses vieses estão presentes na massa de dados utilizada pelos sistemas inteligentes e algorítmicos de aprendizado de máquina. Caso a máquina receba dados e informações carregados de vieses e preconceitos de raça, de gênero, orientação sexual, de forma física ou de qualquer outro traço, ela irá não apenas aprender com eles como perpetuá-los seja durante o seu processo de aprendizado ou quando exposta a novos dados (Garcia, 2020). O problema do viés em sistemas de IA é bastante intensificado pela sua opacidade, portanto, a eliminação do viés requer muitas vezes a análise em conjunto com a opacidade do sistema, pois são dois aspectos que estão intrinsecamente relacionados (Muller, 2020).

Outro aspecto que levanta preocupações diz respeito ao uso indiscriminado de IA para automatizar decisões em esferas sensíveis no qual a interferência humana se mostra essencial, resultando em diferentes formas de discriminação. Na área de saúde, por exemplo, existem preocupações acerca dos efeitos discriminatórios e excludentes dos serviços de saúde baseados em IA, que podem trazer consequências disruptivas para a população em termos de acesso e democratização dos serviços de saúde (Morley *et al*, 2020). Em seu relatório *Ethics and*

Governance of artificial intelligence for Health, a Organização Mundial da Saúde (OMS, 2021), chama atenção para os desafios e riscos do uso da IA na saúde, como coleta antiética de dados, discriminação algorítmica e ameaças à segurança do paciente e privacidade. O uso não regulamentado da IA pode subordinar os direitos dos pacientes aos interesses comerciais das empresas de tecnologia ou aos interesses de governos em vigilância (OMS, 2021).

Na contemporaneidade lidamos com um novo aspecto que aflige os campos da Filosofia da Informação, Computação, Comunicação e BCI reconhecido como a popularização de aplicações baseadas em *Large Language Models* (LLM), isto é, modelos de *machine learning* treinados com grandes quantidades de dados linguísticos e capazes de gerar, compreender e manipular texto de forma avançada usando técnicas de IA, o que resulta em respostas muito precisas. Neste sentido, Li *et al* (2023) levantaram uma série de questões e de incertezas sobre o largo uso dessas ferramentas. A exemplo de questionamentos do uso da inteligência generativa está a popularização do *ChatGPT* em 2023 – aplicação desenvolvida pela empresa *OpenAI*² com função generativa (*e.g.* cria conteúdos diversos como textos, imagens, áudios, vídeos *etc.*) – causou uma repercussão decorrente do uso massivo nos mais diferentes contextos sociais, econômicos e culturais tendo como efeito espontânea o aceleração dos discursos e das pesquisas sobre as implicações éticas decorrentes do seu uso. É possível inferir que o impacto dessa tecnologia no cenário acadêmico transcorre as questões de plágio, uma vez que o sujeito tem possibilidades de criação textual com base nas respostas automáticas da IA, neste caso captadas de fontes *online*, sem designar os devidos créditos de autoria ou mesmo a criação a partir de padrões de comportamento do usuário nas redes, ou seja a criação é feita sem auxílio de fontes informação. Ainda nos casos de produção acadêmica, uso ético da informação e integridade científica é necessário observar os casos em que as IA generativas inventam autorias e documentos pela aglutinação de dados de suas bases para atender às demandas dos atores de IA, neste caso os pesquisadores.

Portanto, a ética em IA é um tópico relevante nos campos das Ciências Sociais Aplicadas, principalmente, para garantir o uso responsável, justo, transparente e equitativo da IA considerando seu impacto nas pessoas, a prevenção de viés e discriminações, a atribuição de responsabilidade, a construção de confiança e a conformidade com normas éticas e legais. Nesta esteira observa-se, o aumento de discursos e pesquisas sobre o tema se traduz no crescente e variado número de publicações que tentam estabelecer parâmetros para o uso ético da IA em termos de responsabilidade e transparência (Burle; Cortiz, 2020).

² Acesso o *OpenAI* em <https://openai.com/>.

3 Procedimentos metodológicos

Este estudo adota uma abordagem quanti-qualitativa com características exploratórias, utilizando a análise bibliométrica e a análise de conteúdo como estratégias complementares. A escolha por essa abordagem mista fundamenta-se na perspectiva de Yin (2016), segundo a qual a integração entre métodos permite a compreensão mais aprofundada de fenômenos complexos, especialmente em campos interdisciplinares como a ética aplicada à Inteligência Artificial (IA).

A análise de conteúdo foi conduzida conforme os princípios clássicos de Bardin (1977), estruturada em três etapas: (1) pré-análise, (2) exploração do material e (3) tratamento dos resultados e interpretação. O objetivo dessa técnica é sistematizar e categorizar o conteúdo textual das publicações de forma a identificar padrões, regularidades, conceitos recorrentes e sentidos latentes. A esse processo, agregaram-se os fundamentos de Bauer e Gaskell (2002), que enfatizam o potencial da análise qualitativa como instrumento de ruptura com a intuição, permitindo a emergência de categorias teóricas a partir dos dados empíricos.

Para esta pesquisa, a base de dados Scopus foi selecionada por duas razões principais: (1) sua ampla cobertura na maioria dos campos de conhecimento, facilidades de acesso e visualização permitem a análise de dados de forma apropriada (Baaset *al.*, 2010; Visser; Van Eck; Waltman, 2021) seu uso em vários estudos bibliométricos (Abad-Segura *et al.*, 2020; Zhu; Liu, 2020). (2) Além disso, a qualidade dos metadados encontrados na Scopus facilitam a exportação para *softwares* de análises bibliométricas como o *VosViewer*.

A coleta de dados foi realizada em março de 2023 utilizando a seguinte estratégia de busca: (TITLE-ABS-KEY (*ethic** AND *artificial* AND *intelligence*AND *plagiarism*) OR TITLE-ABS-KEY (*chatgpt* AND *inteligência* AND *artificial* AND (*segurança* AND *da* AND *informação* OR *information* AND *security*)) OR TITLE-ABS-KEY (“*algoritmos éticos*” OR “*explicabilidade em IA*” OR “*viés em algoritmos*” OR “*privacidade em IA*” OR “*responsabilidade em IA*” OR “*ética em IA*”) AND TITLE-ABS-KEY (“*ethicalalgorithms*” OR “*AI explainability*” OR “*algorithm bias*” OR “*AI privacy*” OR “*AI accountability*” OR “*AI ethics*”)).

Essa estratégia resultou em 504 documentos publicados entre 2015 e 2022, recorte temporal justificado por dois fatores: (i) a intensificação do debate ético sobre IA após a popularização dos modelos generativos e (ii) o crescimento exponencial da produção científica sobre o tema nesse período. Todos os documentos foram exportados com seus metadados completos para análise.

A análise bibliométrica foi utilizada para mapear a evolução da produção científica sobre ética e IA. Os dados foram organizados em planilhas Excel e visualizados no software *VosViewer*, que permitiu gerar mapas de coocorrência de termos e identificação de clusters temáticos.

Foram extraídos indicadores como: evolução anual de publicações, principais países e instituições, periódicos mais relevantes, número de citações e perfil de acesso (aberto ou restrito).

Como critério de seleção, optou-se por analisar os 10 documentos com maior número de citação na base Scopus, pois sugere-se a sua relevância para o entendimento da temática. Alguns aspectos que podem ser analisados a partir do mapeamento realizado incluem: os valores envolvidos, a equidade, os impactos sociais e ambientais, a privacidade, a segurança, a transparência e a responsabilidade (Burle; Cortiz, 2020).

Ampliando os recursos científicos metodológicos da BCI, esta pesquisa também se apoia na análise bibliométrica que é considerada uma alternativa para estudar campos de pesquisa e a sua produção científica (Fahimnia; Sarkis; Davarzani, 2015; Vanti, 2002). Esta metodologia de pesquisa apresenta um processo sistemático, rigoroso e transparente que examina amplamente um campo científico ou uma área de estudo específica (Carrión-Mero *et al.*, 2021; Fahimnia; Sarkis; Davarzani, 2015). Observa-se que para os seus fins, a análise bibliométrica permite o exame do corpus de publicações de modo a evidenciar as relações entre áreas do conhecimento, temáticas, pesquisadores, países e instituições, demonstrando a dinâmica da comunicação científica.

O uso de indicadores bibliométricos pode fornecer *insights* valiosos, quantitativos e qualitativos, sobre diferentes aspectos da pesquisa acadêmica sobre ética e IA (Vanti, 2002). Nesse artigo, inclui-se indicadores relacionados à evolução anual das publicações; os principais países que publicaram sobre o tema; *clusters* de palavras-chave, fontes de publicações sobre ética e IA, fontes das publicações com maior número de citações e o acesso aberto das publicações. Propõe-se que identificar e analisar *clusters*³ pode ajudar a identificar os principais temas, subtemas e tendências de pesquisa em ética e IA fornecendo *insights* sobre os tópicos mais abordados e as conexões entre eles.

Os índices bibliométricos permitem avaliar a produtividade e qualidade das pesquisas acadêmico-científicas, logo a análise da evolução anual das publicações permite observar a tendência de crescimento da produção científica nessa área específica ao longo do tempo (Vanti, 2002). Por exemplo, identificar os anos em que houve aumentos significativos no número de publicações pode indicar o surgimento de novos tópicos de pesquisa ou o aumento do interesse pela ética e IA em determinados períodos.

³ Para melhor compreensão, conceitua-se os clusters de palavras-chave como: grupos de termos que se relacionam semanticamente e são frequentemente utilizados juntos nos artigos científicos.

Além disso, identificar os principais países que publicaram sobre o tema é fundamental para compreender a distribuição geográfica da pesquisa em ética e IA que impactam em fatores e aspectos políticos, econômicos e sociais. Isso ajuda a identificar quais nações estão na vanguarda do desenvolvimento de conhecimentos nessa área e pode indicar a existência de diferentes abordagens culturais ou regulatórias em relação à ética e IA no mundo.

Quanto às fontes de publicações sobre ética e IA, é importante examinar as principais revistas, conferências e periódicos que publicam pesquisas nessa área. Identificar essas fontes de informação é essencial para direcionar futuras pesquisas e para que pesquisadores e profissionais possam acompanhar as últimas descobertas e debates sobre ética e IA. Justifica-se também, as análises realizadas sobre as fontes das publicações com maior número de citações (Vanti, 2002). No cenário de publicações em periódicos científicos é relevante identificar os trabalhos mais influentes e impactantes no campo da ética e IA. Essas fontes são referências importantes e podem fornecer uma visão sobre os estudos que são mais amplamente reconhecidos e citados pela comunidade científica. Vanti (2002, p. 156) ainda ressalta que a bibliometria permite também observar o “[...] uso do número de publicações e citações para auxiliar na avaliação do desempenho científico de pesquisadores, grupos e centros de pesquisa; na tomada de decisões quanto à distribuição de recursos financeiros [...]”.

Por fim, avaliar o acesso aberto das publicações é crucial para garantir a disseminação ampla e o acesso fácil às pesquisas em ética e IA. Identificar a proporção de artigos disponíveis em acesso aberto pode ajudar a compreender a disponibilidade e a acessibilidade da literatura científica nessa área, contribuindo para a transparência e a democratização do conhecimento.

4 Apresentação e discussão dos resultados

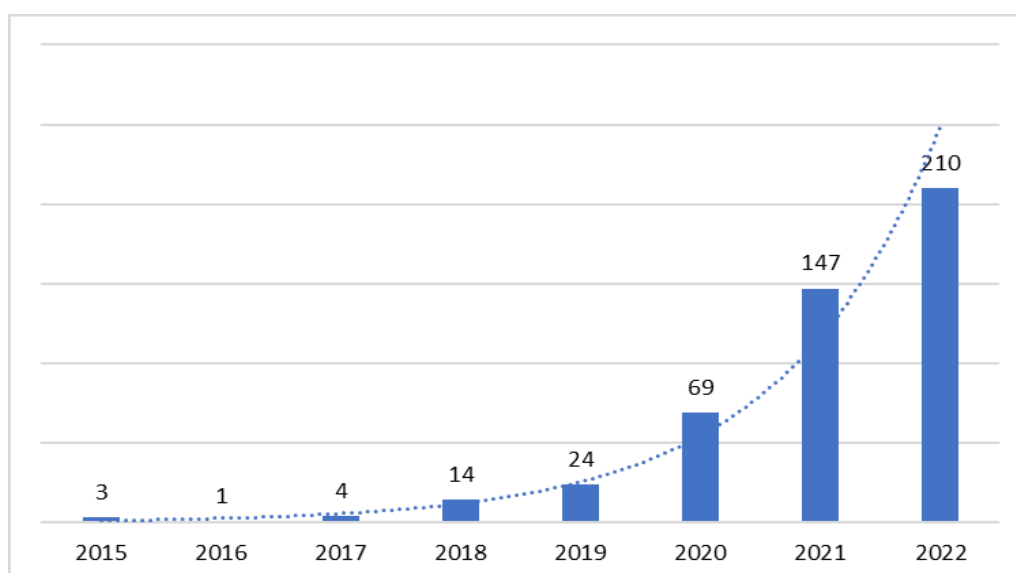
A análise integrada de indicadores bibliométricos (Vanti, 2002) e de conteúdo (Bardin, 1977) permite observar de forma crítica o estado atual da produção científica sobre ética e Inteligência Artificial (IA). A pesquisa combinou o mapeamento quantitativo da literatura internacional com a identificação de padrões conceituais e temáticos, utilizando o software VosViewer para gerar representações visuais das redes de coocorrência de palavras-chave, afiliação de países e fontes de publicação.

4.1 Indicadores bibliométricos da produção científica sobre o tema

De acordo com os dados coletados na Scopus sobre ética e IA, o número de publicações aumentou em 69 vezes entre 2015 e 2022. Esse aumento foi especialmente pronunciado a partir de 2020, quando o número de publicações sobre o tema mais do que dobrou em relação ao ano anterior.

Esses dados sugerem que a publicação sobre ética e IA tem aumentado significativamente nas últimas décadas (Figura 1). À medida que a IA se torna cada vez mais importante do ponto de vista econômico, político, social e cultural. O crescimento de tais produções tende a se manter como um eixo temático contínuo para diversos campos do conhecimento visto a cenário contemporâneo e as questões que envolvem a ética e AI. Do ponto de vista epistemológico, a dimensão ética que orbita os saberes humanos diz respeito ao seu valor moral e de poder, além das dimensões políticas (Schneider, 2013). Além disso, é importante notar que a evolução das publicações sobre ética e IA pode variar em diferentes partes do mundo, dependendo, de mesmo modo, dos fatores culturais, sociais, políticos e econômicos.

Figura 1 - Evolução anual das publicações (2015-2022)⁴



Fonte: dados da pesquisa (2023).

Ao se observar os países de afiliação das autorias que publicam sobre a temática, identificam-se nove *clusters* distintos (Quadro 2):

⁴ Em 1985 recuperou-se a publicação “*IEEE Region 5 Conference*” (1985..., 1985) e em 2008 “*Open source AI*” (Hibbard, 2008). Para calcular a taxa de crescimento entre 2015 e 2022, foi necessário calcular a variação percentual entre o valor inicial (três publicações) e o valor final (210 publicações). Para tal, utilizou-se a fórmula da variação percentual: $\text{variação percentual} = (\text{valor final} - \text{valor inicial}) / \text{valor inicial} \times 100\%$. Substituindo os valores na fórmula, obtêm-se: $\text{variação percentual} = (210 - 3) / 3 \times 100\%$ - variação percentual = 207 / 3 x 100% - variação percentual = 69 x 100% - variação percentual = 6900%.

Quadro 2 – Clusters de países de afiliação das autorias

<i>Cluster</i>	Países	Total de países
Vermelho	China, Finlândia, Jordânia, Malásia, Noruega, Paquistão, Catar, Arábia Saudita e Suécia	Nove
Verde	Austrália, Áustria, Brasil, Canadá, França, Portugal, Singapura, Eslovênia e Vietnã	Nove
Azul	Bulgária, Dinamarca, Grécia, Itália, Malta, México, Polônia e Espanha	Oito
Amarelo	Chipre, Egito, Islândia, Índia, Holanda, África do Sul e Reino Unido	Sete
Lilás	Irã, Macau, Taiwan e Estados Unidos da América (USA)	Quatro
Azul Claro	Alemanha, Gana e Japão	Três
Laranja	República Tcheca e Hungria	Dois
Marrom	Irlanda e Quênia	Dois
Rosa	Suíça	Um

Fonte: elaborado pelas autorias (2023).

As análises de dados da pesquisa apontam que, dentre os países que se destacam pela alta produção científica com a temática ética e IA, os EUA aparece em primeiro lugar com 146 documentos. O país possui várias empresas de tecnologia, universidades e organizações que estão ativamente pesquisando e publicando sobre ética e IA. O Vale do Silício, na Califórnia, sedia algumas das principais empresas que têm liderado a pesquisa e desenvolvimento em IA, tais como *Alphabet*, *Meta*, *Microsoft*, *OpenAI*, entre outras. O Reino Unido (63 documentos) tem sido um líder global em pesquisa sobre ética e IA e tem uma série de organizações importantes, como o *Alan Turing Institute*, que se concentram nesse campo de pesquisa. Vale ressaltar, que o instituto foi fundado para homenagear o cientista da computação Alan Turing (1912-1954)⁵, cujo trabalho pioneiro em matemática teórica e aplicada, engenharia e computação é considerado um dos principais trabalhos de pesquisa que culminaram nos campos da ciência de dados e inteligência artificial. A Finlândia (40 documentos) possui uma sólida economia e uma indústria de tecnologia de ponta. Além disso, o país tem um histórico de forte compromisso com os direitos humanos e a proteção da privacidade. Eles têm adotado uma abordagem ética em relação à inteligência artificial, enfatizando a transparência, a responsabilidade e a equidade.

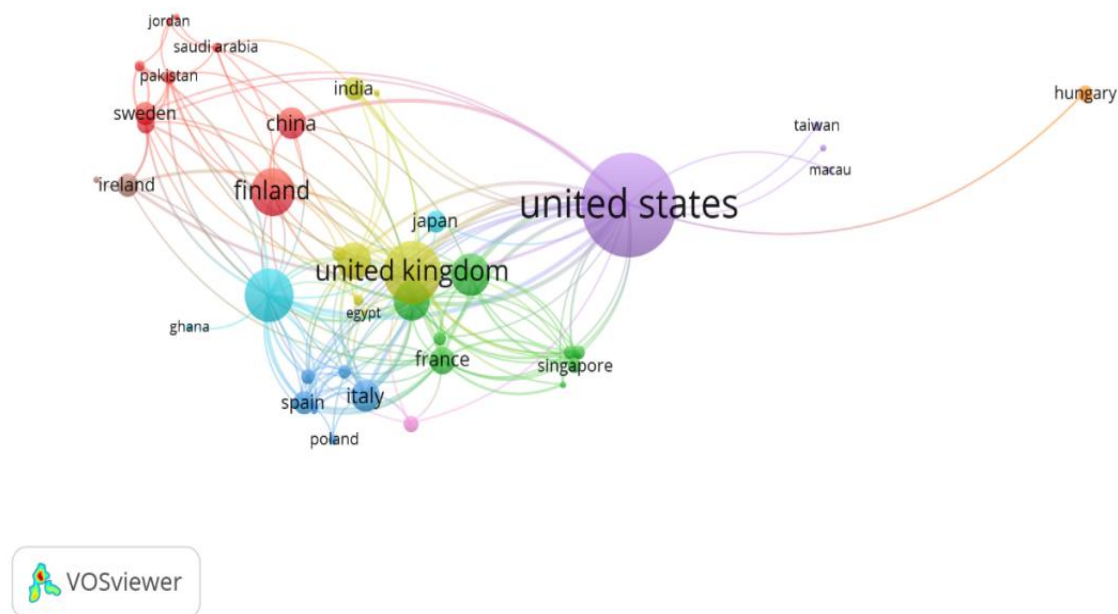
⁵ Alan Turing embora reconhecido herói da Segunda Guerra Mundial e pioneiro na computação pela decifração de códigos, Bletchley Park, foi perseguido pela coroa inglesa por motivos de sua sexualidade. A sua morte por envenenamento com cianureto se deu após a castração química que recebeu por ordem do governo britânico que naquela época que condenava pessoas por homossexualidade.

A Alemanha (49 documentos) tem buscado estabelecer uma estrutura ética sólida para a inteligência artificial. O país está desenvolvendo diretrizes e padrões éticos para governar o uso da IA em áreas como saúde, transporte e indústria.

A China (21 documentos) também está emergindo como um importante centro de pesquisa e desenvolvimento de IA, o ramo empresarial chinês têm trabalhado em questões de ética e IA. Nas questões que tocam o poder das potências globalizadas a criação do *DeepSeeké* um reflexo dos aspectos democráticos globalizados que envolvem economia e política. A França (16 documentos) é um dos principais polos de tecnologia da Europa e, como membro da União Europeia, tem discutido uma série de regulações e diretrizes para uso de tecnologias digitais, especialmente a IA.

Como pode ser observado na Figura 2, esses países não são os únicos envolvidos em pesquisas sobre ética e IA, mas se destacam nos dados analisados. Nota-se que muitos países estão envolvidos no desenvolvimento de políticas e regulamentações para garantir que a IA seja usada de maneira ética e responsável. Para Schneider (2013, p. 61) “a política, portanto, é o território da efetividade da ética – seja de que ética for –, já que a ética sem política é inócua, é potência sem ato”. O que corrobora com a máxima de Cremonese (2019, p. 9): “Quanto mais éticos forem as nossas ações, mais sentido damos a nossa existência”. Portanto o que se espera do comportamento dos sujeitos frente às inovações de IA é o potencial de criticidade no seu uso ético, independente dos meios e suas finalidades.

Figura 2 - Principais países que publicaram sobre o tema



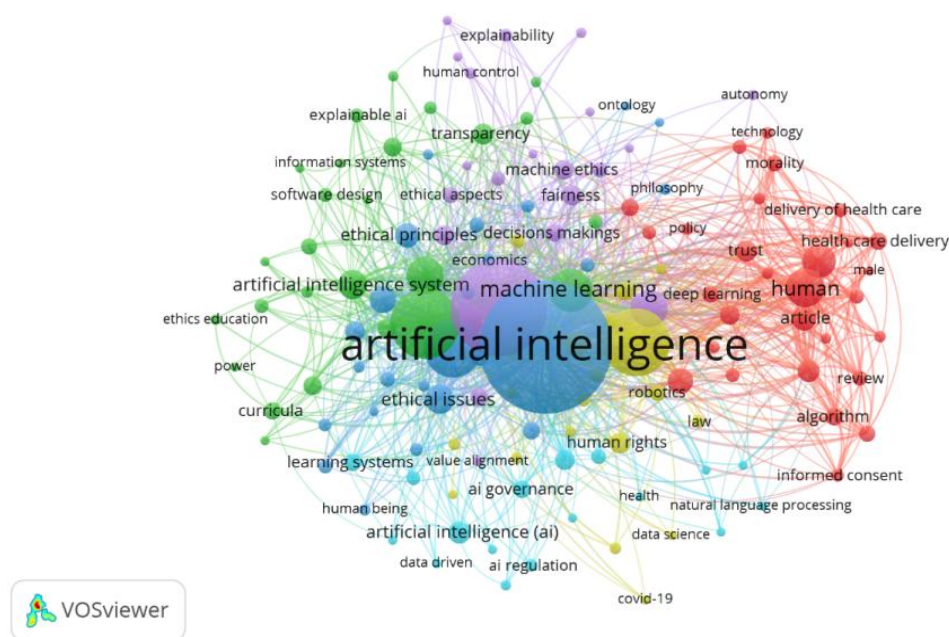
Fonte: dados da pesquisa (2023).

Ainda é necessário pontuar, que segundo Jobin, Ienca e Vayena(2019), os países economicamente mais desenvolvidos estão moldando o debate sobre ética em IA mais do que outros, o que levanta preocupações sobre negligenciar o conhecimento local, o pluralismo cultural e as demandas de justiça global. Assim, os autores buscam demonstrar que as áreas geográficas como África, América do Sul e Central e Ásia Central indica que essas regiões globais não estão participando igualmente do debate sobre a ética da IA evidenciando pouca participação dessas regiões nos debates de ética em IA revelando um desequilíbrio de poder no discurso internacional (Jobin; Ienca; Vayena,2019).

Avançando nas análises dos dados de pesquisa, os *clusters* de palavras-chave (Figura 3, apresentada a seguir) são grupos de termos que estão semanticamente relacionados entre si e são frequentemente utilizados juntos nas publicações sobre ética e inteligência artificial. Desse modo, de um total de 3092 palavras-chave, presentes nas publicações, selecionou-se somente as palavras com no mínimo cinco ocorrências, totalizando 143.

Na Figura 3 observa-se que, as palavras "*Artificial intelligence*" (n=377), "*AI ethics*" (n=225), "*Philosophical aspects*" (n=132), "*Ethics*" (n=124) e "*Ethical Technology*" (n=97) são as mais evidentes e se interconectam. Esses termos estão semanticamente relacionados e são frequentemente utilizados juntos nas publicações mapeadas, sugerindo uma preocupação crescente em abordar as implicações éticas da inteligência artificial. A expressão "*Artificial intelligence*" é amplamente discutida, e seu vínculo com "*AI ethics*" mostra a importância de considerar as questões éticas no desenvolvimento e aplicação da IA. A presença de "*Philosophical aspects*" indica que a discussão sobre ética e IA também envolve reflexões filosóficas sobre os fundamentos morais e conceituais da tecnologia. Isso sugere um interesse em compreender as implicações éticas de um ponto de vista mais amplo. A inclusão do termo "*Ethics*" como uma palavra-chave central reforça a ênfase na importância dos princípios éticos na IA. Esse termo pode abranger diferentes dimensões éticas, como responsabilidade, justiça e privacidade. Por fim, a presença de "*Ethical Technology*" indica o interesse em desenvolver tecnologias que sejam éticas por design, considerando não apenas os aspectos técnicos, mas também os impactos sociais e éticos da IA (Muller, 2020).

Figura 3 - Clusters de palavras-chave⁶



Fonte: dados da pesquisa (2023).

Observa-se que uma análise mais profunda das palavras-chave pode ser útil para identificar os termos e os conceitos mais frequentes nas publicações. Isso pode ajudar a entender como as autorias se referem a diferentes temas e conceitos na literatura sobre ética e IA e quais são os que atendem aos desígnios de seus campos de saber.

Neste ponto, apresentam-se as publicações sobre ética e IA com maior número de citações da Scopus (Quadro 3). A comunicação científica e os seus resultados são observáveis neste extrato em três frentes (Campello; Cendón, Kremer, 2000): a) os **encontros científicos** – em que o(a) cientista assume a posição de comunicador e apresenta propostas de pesquisa, resultados e trabalhos aos pares; b) **anais de eventos** – uma forma intermediária de documentar as pesquisas apresentadas e eventos, essa tipologia documental precede os artigos de periódicos no processo de comunicação científica; c) **periódicos científicos** – meio pelo qual as autorias realizam a divulgação formal de suas pesquisas de forma amplificada, rápida e precisa, além disso os periódicos conservam o conhecimento humano explicitado no formato de artigo.

⁶Para a visualização dinâmica dos resultados, acesse:

<https://app.vosviewer.com/?json=https://drive.google.com/uc?id=1jPqNaBBGrsgqExOpW1gNIVtPBUoldG1T>

Quadro 3 - Fontes de publicações sobre ética e IA

Título	Local	Tipologia do documento
AIES ⁷ 2020 – <i>Proceedings of the AAAI⁸/ACM⁹Conference on AI, Ethics, and Society</i>	USA	Anais
AIES 2021 – <i>Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society</i>	USA	Anais
AIES 2022 – <i>Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society</i>	USA	Anais
<i>Ethics and Information Technology</i>	USA	Periódico
IEEE ¹⁰ <i>Security and Privacy</i>	USA	Periódico
<i>Journal of Business Ethics</i>	Canadá	Periódico
<i>Minds and Machines</i>	USA	Periódico
<i>Proceedings of the ACM on Human-Computer Interaction</i>	USA	Anais
<i>Science and Engineering Ethics</i>	USA	Periódico

Fonte: elaborado pelas autorias (2023).

As publicações de mais destaque na Scopus centralizam os EUA – como foi observado no *cluster* de países com mais publicações –, dentre os anais e periódicos recuperados destaca-se a AAAI/ACM *Conference on AI* (Figura 4) que é um evento anual organizado pela *Association for the Advancement of Artificial Intelligence* em parceria com a *Association for Computing Machinery*¹¹. A conferência tem como objetivo promover a discussão e o debate sobre questões éticas relacionadas à inteligência artificial, incluindo temas como transparência, responsabilidade, justiça e privacidade. O evento reúne especialistas em ética e inteligência artificial de todo o mundo, além de apresentações de artigos científicos sobre as mais recentes pesquisas nessa área. O Canadá está dentre os resultados com o periódico *Journal of Business Ethics*, voltado para as questões éticas relacionadas aos negócios, atendendo ao *Financial Times* como fonte de informação.

⁷Artificial Intelligence, Ethics, and Society (AIES).

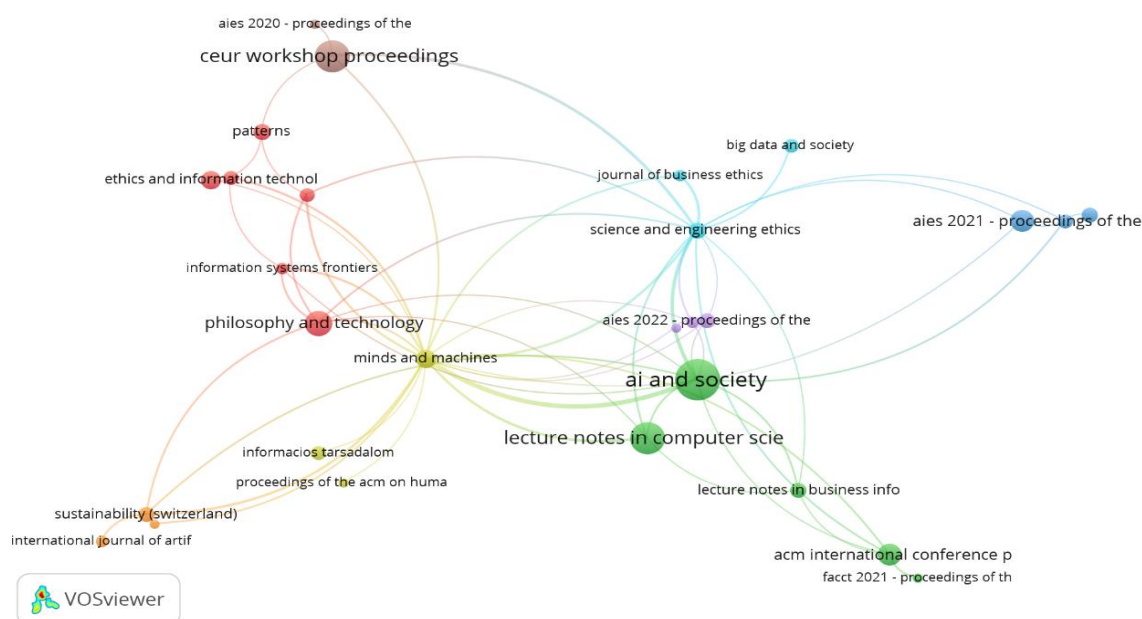
⁸Association for the Advancement of Artificial Intelligence (AAAI).

⁹Association for Computing Machinery (ACM).

¹⁰Institute of Electrical and Electronics Engineers (IEEE)

¹¹Association for Computing Machinery pode ser acessado em: <https://www.aies-conference.com/>

Figura 4 - Fontes das publicações com maior número de citações

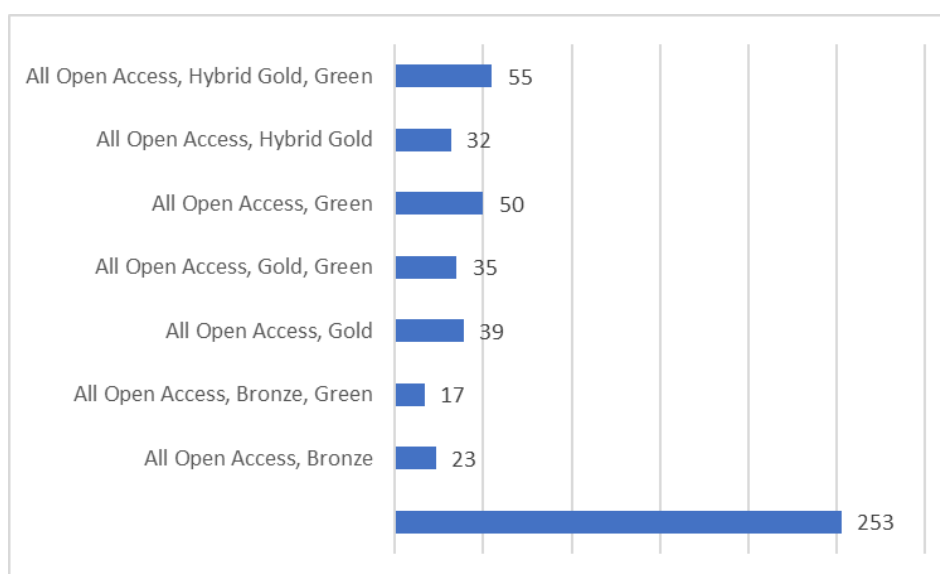


Fonte: dados da pesquisa (2023).

Ao analisar as fontes das publicações, é possível identificar os principais periódicos e conferências em que os pesquisadores publicam seus trabalhos. Isso pode ajudar a entender quais são as principais plataformas para se obter informações atualizadas e confiáveis sobre o tema pesquisado.

No que diz respeito às publicações em acesso aberto, a Figura 5 apresenta a distribuição dos tipos de acesso dos artigos científicos identificados na análise bibliométrica sobre ética e Inteligência Artificial. A categoria "Acesso restrito" (n=253) corresponde aos artigos que só podem ser consultados mediante assinatura ou pagamento. Já a classificação "All Open Access" (n=251) engloba os artigos disponibilizados gratuitamente na internet, embora possam existir diferentes níveis de restrição quanto ao uso e compartilhamento do conteúdo.

Figura 5 – Distribuição dos tipos de acesso às publicações científicas



Fonte: elaborado pelas autorias (2023).

As categorias "*Gold*" (n=39) indicam que o artigo foi publicado em uma revista de acesso aberto e o autor pagou uma taxa para disponibilizá-lo, enquanto a categoria "*Green*" (n=50) indica que o autor publicou o artigo em um repositório de acesso aberto. A categoria "*Hybrid Gold*" (n=32) indica que o artigo foi publicado em uma revista que oferece tanto acesso aberto quanto acesso restrito, e o autor optou por torná-lo disponível gratuitamente. A análise desses dados pode ajudar a entender como os pesquisadores escolhem as revistas em que publicam seus artigos e como isso afeta o acesso à pesquisa sobre ética e IA.

É importante salientar que o acesso aberto, pois ele permite que a produção acadêmica e científica seja amplamente disseminada e acessível a todas as pessoas, independentemente de sua afiliação institucional ou capacidade de pagar por acesso. Isso aumenta a visibilidade e o impacto da pesquisa, promove a colaboração e a inovação, além de ser uma questão de equidade e justiça social. Além disso, o acesso aberto é fundamental para acelerar o progresso científico e melhorar a qualidade de vida das pessoas ao redor do mundo, especialmente em países em desenvolvimento que muitas vezes não têm acesso aos recursos necessários para pagar por acesso a publicações científicas e acadêmicas.

4.2 Análise de conteúdo das publicações mais citadas

Como demonstrado teoricamente, o campo da IA está evoluindo rapidamente e à medida que se torna cada vez mais integrado em nossas vidas diárias, há uma necessidade crescente de considerar as implicações éticas da IA. Com base na análise sobre o conteúdo dos 10 documentos mais citados na base Scopus (Quadro 4), é possível sintetizar os resultados e

identificar as principais tendências e temas na literatura sobre ética e IA que atendem aos propósitos desta pesquisa.

Quadro 4 - Análise de conteúdo das 10 publicações mais citadas

Publicações mais citadas	Ano de publicação	Total de citações
<i>The ethics of AI Ethics: an evaluation of guidelines</i>	2020	333
<i>From what to how: an initial review of publicly available AI ethics tools, methods, and research to translate principles into practices</i>	2020	154
<i>Global evolution of research in artificial intelligence in health and medicine: a bibliometric study</i>	2019	143
<i>Artificial Intelligence and surgical decision-making</i>	2020	110
<i>Ethics and privacy in AI and Big Data: implementing responsible research and innovation</i>	2018	77
<i>Algorithmic injustice: a relational ethics approach</i>	2021	76
<i>Ethical considerations in Artificial Intelligence courses</i>	2017	68
<i>The role and limits of principles in AI ethics: towards a focus on tensions</i>	2019	68
<i>A misdirected principle with a catch: explicability for AI</i>	2019	58
<i>Artificial Intelligence (AI) ethics: ethics of AI and ethical AI</i>	2020	58

Fonte: elaborado pelas autorias (2022).

O artigo “*The ethics of AI ethics: an evaluation of guidelines*” publicado em 2020 na revista *Minds and Machines*, recebeu 333 citações. O artigo ressalta que a IA tem sido um dos avanços tecnológicos mais significativos do século XXI, revolucionando a forma como vivemos e trabalhamos. Os sistemas de IA têm o potencial de fornecer soluções para alguns dos desafios mais urgentes do mundo, como saúde, mudança climática e desigualdade social. No entanto, com esse potencial vem um conjunto de desafios éticos que devem ser enfrentados para garantir o desenvolvimento e uso seguro e responsável de sistemas de IA (Hagendorff, 2020).

Em resposta a essas preocupações, várias organizações e iniciativas desenvolveram diretrizes para o desenvolvimento e uso ético da IA. Essas diretrizes geralmente descrevem princípios e recomendações para garantir que os sistemas de IA sejam desenvolvidos e usados de maneira ética, transparente e responsável. Assim, Hagendorff (2020) analisou e comparou as diretrizes para o desenvolvimento e uso ético da IA para fornecer uma visão abrangente do campo da ética da IA e destacar as áreas em que é necessária mais atenção. Para tanto, conduziu uma avaliação sistemática de 22 diretrizes éticas de IA incluindo aquelas desenvolvidas por governos, organizações internacionais, associações industriais e instituições acadêmicas. As diretrizes foram identificadas por meio de uma revisão da literatura e foram avaliadas com base em seu escopo, princípios e recomendações.

A análise revelou várias áreas de sobreposição entre as diretrizes. A maioria das diretrizes enfatizou a importância da transparência, responsabilidade e justiça no desenvolvimento e uso de sistemas de IA. Também houve consenso sobre a necessidade de supervisão e responsabilidade humana na tomada de decisões, bem como a importância de respeitar a privacidade e a proteção de dados. (Hagendorff, 2020).

A análise também revelou algumas omissões nas diretrizes, como a falta de ênfase no impacto dos sistemas de IA no emprego e nos direitos trabalhistas, bem como falta de atenção ao impacto ambiental dos sistemas de IA. Hagendorff (2020) destaca que para garantir que as diretrizes de ética da IA sejam eficazes, é importante considerar como elas podem ser implementadas na prática. Isso inclui desenvolver ferramentas e métodos para avaliar as implicações éticas dos sistemas de IA, bem como fornecer treinamento e educação para pesquisadores, desenvolvedores e usuários de sistemas de IA. Além disso, há necessidade de maior colaboração e coordenação entre as partes interessadas no desenvolvimento e uso de sistemas de IA incluindo governos, indústria, academia e sociedade civil.

A ética da IA é um campo importante e em rápida evolução, com um número crescente de organizações e iniciativas desenvolvendo diretrizes para o desenvolvimento e uso ético de sistemas de IA. Vales ressaltar, que o desenvolvimento responsável e o uso de sistemas de IA exigirão atenção contínua às questões éticas e o compromisso de defender os princípios éticos diante das mudanças tecnológicas (Hagendorff, 2020).

A publicação "*From what to how: an initial review of publicly available AI ethics tools, methods and research to translate principles into practices*" recebeu 154 citações. O objetivo do texto foi investigar as lacunas entre os princípios éticos da IA e as práticas reais dos desenvolvedores de IA, principalmente na área de *Machine Learning*. O artigo apresenta uma metodologia para ajudar os desenvolvedores a aplicarem a ética em cada estágio do *pipeline* de desenvolvimento do *Machine Learning* e identifica a necessidade de abordagens mais coordenadas e sofisticadas para traduzir os princípios éticos em protocolos de design (Morley et al., 2020).

O artigo também destaca a distribuição desigual de esforços na tipologia "*Applied AI Ethics*" e a imaturidade de muitas das ferramentas incluídas, o que pode dificultar o incentivo de sua adoção por desenvolvedores de *Machine Learning*. A falta de orientação e incentivos para desenvolver IA de forma ética pode resultar em custos de erros éticos superando os benefícios dos recursos de AI, prejudicando a aceitação pública e potencialmente levando a uma perda de confiança de investidores e financiadores de pesquisa de IA (Morley et al., 2020).

O terceiro artigo "*Global evolution of research in artificial intelligence in health and medicine: a bibliometric study*" que recebeu 143 citações, discute o uso crescente da IA na saúde

e na medicina e fornece uma visão histórica e global da pesquisa neste campo. O estudo utilizou técnicas bibliométricas e analisou 27.451 artigos publicados entre 1977 e 2018 e descobriu que a maioria foi publicada entre 2008 e 2018. O estudo examinou o volume de publicação, a colaboração entre autorias e países e identificou as principais técnicas e suas aplicações na previsão e tratamento clínico. As autorias constataram que o maior número de publicações estava relacionado ao câncer, seguido por doenças cardíacas e acidente vascular cerebral, deficiência visual, Alzheimer e depressão. O texto sugere que há escassez de pesquisas sobre a aplicação da IA em algumas doenças de alta carga e pede o desenvolvimento de protocolos e regulamentações globais e nacionais sobre a justificativa e adaptação de produtos médicos de IA (Tran *et al.*, 2019).

O artigo "*Artificial Intelligence and surgical decision-making*" discute o uso potencial da IA na tomada de decisões cirúrgicas. As autorias argumentam que a IA pode fornecer informações e suporte valiosos para os cirurgiões, principalmente em casos complexos. Fornece uma visão geral da IA e suas aplicações atuais na área da saúde, incluindo análise de imagens, processamento de linguagem natural e modelagem preditiva. As autorias então se concentram especificamente em como a IA pode ser usada na tomada de decisões cirúrgicas (Loftus *et al.*, 2020).

As autorias argumentam que a IA pode ajudar os cirurgiões a identificarem o melhor curso de ação, analisando uma ampla gama de dados, incluindo histórico do paciente, resultados de laboratório, imagens e outras informações clínicas. Isso pode ajudar os cirurgiões a tomar decisões mais informadas e reduzir o risco de complicações. Também discutem os desafios e limitações da IA na tomada de decisões cirúrgicas, incluindo a necessidade de dados de alta qualidade, o potencial de viés e a importância de manter a supervisão humana (Loftus *et al.*, 2020).

Em "*Ethics and privacy in AI and Big Data: implementing responsible research and innovation*", Stahl e Wright (2018) discutem as considerações éticas e de privacidade relacionadas ao uso de IA e *big data*. As autorias argumentam que, embora a IA e o *big data* tenham o potencial de revolucionar muitos campos, incluindo saúde, finanças e transporte, seu uso também levanta importantes questões éticas e de privacidade.

Exploram o conceito de pesquisa e inovação responsáveis, que enfatiza a necessidade de o desenvolvimento da ciência e da tecnologia levar em conta as necessidades e os valores da sociedade como um todo. As autorias então discutem as considerações éticas relacionadas ao uso de IA e *big data*, incluindo questões de viés, transparência e responsabilidade (Stahl; Wright, 2018).

As preocupações sobre a privacidade são levantadas pelo uso de IA e *big data*, incluindo o potencial de identificação e rastreamento de indivíduos sem seu consentimento, bem como o potencial de uso indevido ou vazamento de informações confidenciais. As autorias argumentam que essas preocupações com a privacidade devem ser levadas a sério e que salvaguardas e regulamentações apropriadas devem ser implementadas para proteger a privacidade dos indivíduos (Stahl; Wright, 2018).

“Algorithmic injustice: a relational ethics approach”, de Birhane (2021), discute o problema do viés algorítmico e seu impacto nas comunidades marginalizadas. A autora argumenta que as abordagens atuais sobre viés algorítmico, como soluções técnicas e estruturas legais, falham em abordar adequadamente as causas do problema.

Em vez disso, Birhane (2021) propõe uma abordagem de ética relacional à justiça algorítmica que enfatiza a importância de entender e abordar o contexto social e histórico da tomada de decisão algorítmica. Essa abordagem exige o reconhecimento dos fatores complexos e inter-relacionados que contribuem para o viés algorítmico, incluindo dinâmicas de poder, injustiças históricas e desigualdades sociais.

A autora argumenta que uma abordagem de ética relacional à justiça algorítmica requer uma mudança na forma como abordamos o desenvolvimento e a implantação da tecnologia. Isso inclui um foco no envolvimento da comunidade, colaboração interdisciplinar e reflexão crítica contínua sobre as implicações éticas dos sistemas algorítmicos (Birhane, 2021).

O artigo *“Ethical considerations in Artificial Intelligence courses”* de Burton e outros (2019) discute a importância de incorporar considerações éticas em cursos de IA. As autorias argumentam que a educação em IA não deve se concentrar apenas em habilidades técnicas, mas também na tomada de decisões éticas, especialmente devido ao potencial da IA de ter um impacto social significativo.

O artigo discute vários tópicos éticos que podem ser incorporados aos cursos de IA, como justiça, transparência, privacidade e preconceito. As autorias fornecem exemplos de como esses tópicos podem ser ensinados em sala de aula, como por meio de estudos de caso ou discussões em grupo (Burton *et al.*, 2019).

Enfatizam a importância da colaboração interdisciplinar na educação em IA observando que as considerações éticas exigem contribuições de especialistas em vários campos, incluindo filosofia, direito e ciências sociais. Sugerem que os cursos de IA podem se beneficiar de palestras ou projetos colaborativos que envolvam especialistas dessas áreas. A incorporação de considerações éticas nos cursos de IA pode ajudar a garantir que os futuros profissionais de IA estejam preparados para tomar decisões responsáveis sobre o desenvolvimento e o uso da tecnologia de IA (Burton *et al.*, 2019).

No artigo *"The role and limits of principles in AI ethics: towards a focus on tensions"*, Whittlestone e outros (2019) argumentam que as abordagens atuais para a ética da IA, baseadas no uso de princípios éticos, são limitadas em sua capacidade de orientar o desenvolvimento e o uso da IA na prática. Eles sugerem que uma abordagem mais produtiva para a ética da IA seria focar nas tensões que surgem em contextos específicos de desenvolvimento e uso da IA. Seria relevante desenvolver abordagens para resolver essas tensões de uma forma que respeite os valores e interesses de todas as partes interessadas.

Foi observado que os princípios éticos são muito abstratos e gerais para fornecer orientação prática para o desenvolvimento e uso da IA (Whittlestone *et al.*, 2019). Eles destacam que os princípios geralmente estão abertos à interpretação e que diferentes partes interessadas podem ter diferentes interpretações do mesmo princípio. Além disso, os princípios costumam estar em tensão uns com os outros e pode ser difícil determinar qual princípio deve ter precedência em um determinado contexto.

Whittlestone e outros (2019) propõem uma abordagem alternativa para a ética da IA com base no conceito de "tensões". Eles argumentam que as tensões surgem quando há interesses, valores ou objetivos conflitantes entre diferentes partes interessadas em um contexto particular de desenvolvimento e uso de IA. Eles sugerem que identificar e abordar essas tensões é uma maneira mais produtiva de orientar o desenvolvimento e o uso da IA do que confiar em princípios éticos.

Dentre as principais tensões destacam-se a privacidade e segurança, transparência e confidencialidade e a autonomia e responsabilidade. Eles argumentam que resolver essas tensões requer uma abordagem contextual e colaborativa que envolve todas as partes interessadas no desenvolvimento e uso de sistemas de IA.

A publicação *"A misdirected principle with a catch: explicability for AI"* de Robbins (2019), argumenta que o princípio da "explicabilidade" na ética da IA pode não ser eficaz para abordar as preocupações em torno da tomada de decisões da IA. O artigo argumenta que, embora a explicabilidade possa parecer uma solução direta para as preocupações relacionadas à tomada de decisões da IA, na verdade é um princípio mal direcionado porque falha em abordar as causas básicas dessas preocupações.

Sugere que a explicabilidade pode não ser eficaz para abordar as preocupações em torno da tomada de decisões da IA por vários motivos. Em primeiro lugar, nem sempre é possível fornecer uma explicação completa de como um sistema de IA chegou a uma decisão específica, principalmente para sistemas complexos. Em segundo lugar, mesmo que seja fornecida uma explicação, ela pode não ser compreensível e significativa para a pessoa que a procura. Finalmente, mesmo que uma explicação seja compreensível, ela pode não abordar

necessariamente as preocupações subjacentes sobre justiça, preconceito e discriminação (Robbins, 2019).

Assim, em vez de confiar apenas na explicabilidade como solução para as preocupações relacionadas à tomada de decisões da IA, uma abordagem mais eficaz pode ser focar nos princípios e valores éticos subjacentes que orientam o desenvolvimento e o uso da IA. Isso pode envolver a consideração de questões sobre quem é responsável pelas decisões tomadas pelos sistemas de IA, como garantir justiça e equidade na tomada de decisões da IA e como abordar questões de preconceito e discriminação (Robbins, 2019).

O artigo "*Artificial Intelligence (AI) ethics: ethics of AI and ethical AI*" de Siau e Wang (2020), explora o conceito de ética da IA e a necessidade de uma IA ética. Definem a ética da IA como "o estudo das questões morais e éticas que envolvem o desenvolvimento e a implantação de máquinas inteligentes" (Siau; Wang, 2020, p. 75, tradução nossa). Eles observam que a ética da IA não é um campo novo, mas sua importância aumentou nos últimos anos devido ao rápido desenvolvimento da tecnologia de IA e sua integração na sociedade. As autorias apontam que existem várias preocupações éticas em torno do desenvolvimento e uso da IA. A pesquisa sobre questões éticas da IA se enquadra em três categorias: características da IA que podem dar origem a problemas éticos, fatores humanos que causam riscos éticos e impacto social de questões éticas da IA. Alguns exemplos dessas preocupações incluem o potencial da IA perpetuar preconceitos e discriminação, o impacto no emprego e no deslocamento de empregos, segurança de dados e preocupações com a privacidade e o potencial da IA ser usada de maneira prejudicial se não for desenvolvida e usada de forma ética. Para obter mais informações sobre segurança de dados e questões de privacidade especificamente. Exploram várias questões éticas importantes relacionadas à IA incluindo privacidade, transparência, responsabilidade, preconceito e o impacto potencial da IA nos empregos e na sociedade. Siau e Wang (2020) argumentam que é essencial abordar essas questões para garantir que a IA seja desenvolvida e implantada de maneira ética e responsável.

Siau e Wang (2020), argumentam que a IA ética é essencial para construir confiança na IA e garantir que ela seja usada em benefício da sociedade. Concluem o artigo ressaltando a importância de um esforço colaborativo entre as partes interessadas, incluindo governo, indústria, academia e sociedade civil, para desenvolver diretrizes e padrões éticos para a IA. Enfatizam também a importância da pesquisa e discussão contínuas nessa área para garantir que a IA seja desenvolvida e usada de forma a promover o bem público e se alinhar aos princípios éticos.

4.3 Lacuna científica nas produções mapeadas

Diante dessas análises, o que se observa entre os documentos mais citados é que o uso de IA para escrever trabalhos acadêmicos, não foi abordado em nenhuma das publicações mais citadas levantadas por este estudo. No entanto, é importante levantar várias questões éticas relacionadas ao uso da IA para escrever trabalhos acadêmicos, tais como:

- Plágio: sistemas de IA podem gerar conteúdo que não é original, e isso pode ser percebido como desonestidade acadêmica.
- Deturpação: os sistemas de IA podem produzir artigos que não são escritos pelas autorias, levando a uma situação em que outra pessoa assume o crédito pelo trabalho.
- Viés: os sistemas de IA são tão bons quanto os dados nos quais são treinados e, se os dados forem tendenciosos, o sistema também produzirá resultados de pesquisa tendenciosos.
- Controle de qualidade: os sistemas de IA podem não produzir trabalhos de alta qualidade e isso pode levar a uma situação em que os alunos sejam avaliados em trabalhos abaixo do padrão.
- Responsabilidade: pode ser difícil determinar quem é o responsável quando algo dá errado com um sistema de IA, e isso pode levar a dilemas legais e éticos.

Como dito, o uso de IA para escrever trabalhos acadêmicos levanta questões sobre a integridade científica, plágio, deturpação, viés, controle de qualidade e responsabilidade. Sobretudo, é importante ressaltar que todas as decisões são políticas e estas causam impacto sobre sujeitos e coletivos de maneiras diversas. A IA não é uma entidade cognitiva superior dotada de lógica computacional sem falhas (Moraes, 2019), ainda que ela atinja o grau de superinteligência (Muller, 2020), uma vez que o seu *fazer* está atrelado à ação humana. Portanto, é imprescindível deixar de ressaltar que a supervisão humana é o ponto crítico essencial para a ética em IA (OCDE, 2019) seja na utilização de *prompts* nas IA generativas ou no uso de *softwares* para as tomadas de decisão.

5 Considerações finais

A Inteligência Artificial (IA) emergiu como uma tecnologia disruptiva, capaz de impulsionar inovações em diversas áreas do conhecimento, promovendo avanços significativos no campo econômico, político e social. No entanto, seu crescimento acelerado também suscita desafios éticos complexos, que precisam ser abordados de forma sistemática para garantir seu desenvolvimento responsável. O presente estudo contribuiu para a compreensão desse cenário

ao mapear e analisar a produção científica sobre ética e IA na base Scopus, evidenciando tendências e lacunas na literatura acadêmica.

Os resultados demonstram que o debate sobre ética e IA tem se intensificado, especialmente a partir de 2020, refletindo uma crescente preocupação com a transparência, responsabilidade, privacidade, vieses algorítmicos e implicações sociais da IA. A análise bibliométrica revelou que países como Estados Unidos, Reino Unido, Alemanha, Finlândia e China lideram a produção acadêmica na área, sugerindo um desequilíbrio no discurso global sobre a ética da IA, com sub-representação de regiões como África, América Latina e Ásia Central.

A revisão de conteúdo das publicações mais citadas na Scopus reforça a centralidade de questões como justiça algorítmica, responsabilidade e explanação de decisões automatizadas. Embora os debates avançados sobre ética em IA sejam amplos e interdisciplinares, observa-se uma lacuna significativa no tocante às implicações do uso de IA para a elaboração de trabalhos acadêmicos, como os riscos de plágio, deturpação de autoria, viés nas fontes utilizadas e a responsabilidade pelo conteúdo gerado por ferramentas de IA generativa.

Os modelos generativos como o ChatGPT têm sido largamente utilizados em contextos acadêmicos e profissionais. No meio acadêmico, o uso generalizado dessa ferramenta tem gerado muitas incertezas sobre questões de transparência, proteção de dados, autoria, originalidade e vieses. Os vários desafios colocados colocam a urgência de regulamentações e de diretrizes claras para garantir a integridade científica, que são entendidas como honestidade, confiabilidade, transparência e responsabilidade. As ações contrárias e danosas perpassam o plágio, a invenção de autorias e de documentos, a aglutinação de dados, a deturpação de conceitos e os vieses algorítmicos. O presente estudo enfatiza que a IA, apesar de suas capacidades avançadas, não é um sistema autônomo sem falhas ou neutralidade, mas sim um reflexo das intenções e valores de seus desenvolvedores e usuários. Assim, o controle humano deve permanecer como elemento essencial na utilização da IA, seja na análise de dados, na tomada de decisões ou na elaboração de conteúdo acadêmico. Mais estudos sobre os impactos do uso da IA Generativa na pesquisa acadêmica, bem como sobre quais as soluções cabíveis para estas práticas quando danosas são necessários dada a sua novidade e complexidade.

Por fim, ao destacar as principais discussões sobre ética e IA e evidenciar lacunas no debate acadêmico, este estudo reforça a necessidade de uma abordagem transdisciplinar na construção de diretrizes éticas e políticas para a IA. A promoção de pesquisas que integrem perspectivas filosóficas, jurídicas, sociais e tecnológicas é fundamental para garantir um desenvolvimento equitativo e transparente dessa tecnologia, assegurando que ela beneficie a sociedade de maneira justa e responsável.

Agradecimentos

Ao Núcleo de Estudos e Pesquisas sobre Recursos, Serviços e Práxis Informacionais (NERSI) da Universidade Federal de Minas Gerais (UFMG), em especial aos membros da linha “Uso ético da Informação” pelo avanço nos debates e reflexões acadêmicas e profissionais.

Referências

- ABAD-SEGURA, E. *et al.* Effects of circular economic policies on the environment and sustainable growth: worldwide research. **Sustainability**, [s.l.], v. 12, n. 14, p. 5792, 2020. Disponível em: <https://www.mdpi.com/2071-1050/12/14/5792>. Acesso em: 01 maio 2023.
- AI ETHICS LAB. **About**. Cambridge: AI EthicsLab, 2022. Disponível em: <https://aiethicslab.com/about/>. Acesso em: 05 maio 2023.
- AI NOW INSTITUTE. **Home**. [s.l.]: AI NowInstitute, 2023. Disponível em: <https://ainowinstitute.org/>. Acesso em: 05 maio 2023.
- ARBIX, G. A transparência no centro da construção de uma ia ética. **Novos estudos CEBRAP**, São Paulo, v. 39, n. 2, p. 395-413, 2020. Disponível em: <https://www.scielo.br/j/nec/a/pD9k5gtHpXwsgFcsMC5gbJg/?format=pdf&lang=pt>. Acesso em: 01 maio 2023.
- ARTIFICIAL INTELLIGENCE. *In*: CAMBRIDGE DICTIONARY. Cambridge: Cambridge University Press & Assessment, 2023. Disponível em: <https://dictionary.cambridge.org/dictionary/english/artificial-intelligence>. Acesso em: 05 maio 2023.
- ASSANGE, J. **Quando o Google encontrou o WikiLeaks**. São Paulo: Boitempo, 2015.
- BARDIN, L. **Análise de conteúdo**. São Paulo: Edições 70, 1977.
- BASS, M. S. *et al.* Global conservation significance of Ecuador’s Yasuní National Park. **PLoS ONE**, [s.l.], v. 5, [s.n.], 2010. DOI <https://doi.org/10.1371/journal.pone.0008767>.
- BIRHANE, A. Algorithmic injustice: a relational ethics approach. **Patterns**, [s.l.], v. 2, n. 2, 2021. DOI: <https://doi.org/10.1016/j.patter.2021.100205>.
- BURLE, C.; CORTIZ, D. **Mapeamento de princípios de inteligência artificial**. São Paulo: Núcleo de Informação e Coordenação do Ponto BR, 2020. Disponível: https://nic.br/media/docs/publicacoes/17/20200721143359/digital_mapeamento_principios_IA_portugues.pdf. Acesso em: 17 maio 2023.
- BURTON, E. *et al.* Ethical Considerations in Artificial Intelligence Courses. **AI Magazine**, Palo Alto, v. 38, n. 2, p. 22-34, summer 2017. Disponível em: <https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/2731>. Acesso em: 05 maio 2023.
- CAMPELLO, B. S.; CENDÓN, B. V.; KREMER, J. M. **Fontes de informação para pesquisadores e profissionais**. Belo Horizonte: Editora UFMG, 2000.

CARRIÓN-MERO, P. *et al.* Worldwide Research Trends in Landslide Science. **Int. J. Environ. Res. Public Health**, [s.l.], v. 18, n. 18, 2021. DOI <https://doi.org/10.3390/ijerph18189445>.

COMISSÃO EUROPEIA. **Ethics guidelines for trustworthy AI**. Bruxelas: European Commission, 2019. Disponível em: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>. Acesso em: 05 maio 2023.

CORTIZ, D. Inteligência artificial: equidade, justiça e consequências. **Panorama Setorial da Internet**, São Paulo, v. 12, n. 1, 2020. Disponível: <https://cetic.br/pt/publicacao/ano-xii-n-1-inteligencia-artificial-equidade-justica-e-consequencias/>. Acesso em: 05 maio 2023.

CREMONESE, D. Ética e moral na contemporaneidade. **Campos Neutrais**: Revista Latino-Americana de Relações Internacionais, Rio Grande, v. 1, n. 1, p. 8-28, jan./abr. 2019. DOI 10.14295/cn.v1i1.8618.

DANAHER, J. The threat of algocracy: reality, resistance and accommodation. **Philosophy & Technology**, [s.l.], v. 29, n. 3, p. 245-268, 2016. Disponível em: <https://link.springer.com/article/10.1007/s13347-015-0211-1>. Acesso em: 05 maio 2023.

FAHIMNIA, B.; SARKIS, J.; DAVARZANI, H. Green supply chain management: review and bibliometric analysis. **Int. J. Prod. Econ.**, [s.l.], v. 162, [s.n.], p. 101–114, Apr. 2015. DOI <https://doi.org/10.1016/j.ijpe.2015.01.003>.

GARCIA, A. C. Ética e Inteligência Artificial. **Computação Brasil**, [s.l.], [s.n.], n. 43, p. 14–22, nov. 2020. Disponível em: <https://sol.sbc.org.br/journals/index.php/comp-br/article/view/1791/1625>. Acesso em: 17 maio 2023.

HAGENDORFF, T. The ethics of AI ethics: an evaluation of guidelines. **Minds and Machines**, [s.l.], v. 30, n. 1, p. 99–120, 2020. Disponível em: <https://link.springer.com/article/10.1007/s11023-020-09517-8>. Acesso em: 21 abr. 2023.

HIBBARD, Bill. Open source AI. **Frontiers in Artificial Intelligence and Applications**, [s.l.], v. 171, n. 1, p. 473-477, 2008. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84875941203&partnerID=40&md5=9bfa745a75e9a3658c8fd2a2b50d6093>. Acesso em: 19 mar. 2023.

HUMAN-CENTERED ARTIFICIAL INTELLIGENCE [HAI]. **Home**. Stanford: Stanford University, 2021. Disponível em: <https://hai.stanford.edu/>. Acesso em: 07 maio 2023.

JASANOFF, S. **States of knowledge**: the co-production of science and social order. London: Routledge, 2004.

JOBIN, A.; IENCA, M.; VAYENA, E. The global landscape of AI ethics guidelines. **Nature Machine Intelligence**, [s.l.], v. 1, n. 9, p. 389–399, set. 2019. Disponível em: <https://www.nature.com/articles/s42256-019-0088-2>. Acesso em: 07 maio 2023.

LATOUR, B. **Pandora's hope**: essays on the reality of science studies. Cambridge: Harvard University Press, 1999.

LI, H. *et al.* Ethics of large language models in medicine and medical research. **The Lancet Digital Health**, [s.l.], v. 5, n. 6, p. 333–335, abr. 2023. DOI [https://doi.org/10.1016/S2589-7500\(23\)00083-3](https://doi.org/10.1016/S2589-7500(23)00083-3).

LOFTUS, T. J. *et al.* Artificial Intelligence and surgical decision-making, **JAMA Surgery**, Chicago, v. 155, n. 02, p. 148-158, fev. 2020. DOI [10.1001/jamasurg.2019.4917](https://doi.org/10.1001/jamasurg.2019.4917).

MORAES, J. A. **O paradigma da complexidade e a Ética Informacional**. Campinas: CLE, 2019. Disponível em: <https://www.cle.unicamp.br/ebooks/index.php/publicacoes/catalog/book/7>. Acesso em: 12 fev. 2025.

MORLEY, J. *et al.* From what to how: an initial review of publicly available AI ethics tools, methods and research to translate principles into practices. **Science and Engineering Ethics**, Califórnia, v. 26, n. 04, p. 2141-2168, ago. 2020. DOI [10.1007/s11948-019-00165-5](https://doi.org/10.1007/s11948-019-00165-5).

MULLER, V. C. **Ethics of Artificial Intelligence and Robotics**. Stanford: Stanford Encyclopedia of Philosophy, 2020. Disponível em: https://plato.stanford.edu/entries/ethics-ai/?utm_source=summari. Acesso em: 05 maio 2023.

ORGANIZAÇÃO MUNDIAL DA SAÚDE [OMS]. **Ethics and governance of artificial intelligence for health: WHO guidance**. [s.l.]: World Health Organization, 2021. Disponível em: <https://apps.who.int/iris/bitstream/handle/10665/341996/9789240029200-eng.pdf>. Acesso em: 20 mai. 2023.

ORGANIZAÇÃO PARA COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO [OCDE]. **Recommendation of the Council on Artificial Intelligence**. Paris: OCDE, 2019. Disponível em: <https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449>. Acesso em: 05 mai. 2023.

PARTNERSHIP ON AI [PAI]. **About us**. [s.l.]: Partnership on AI, 2023. Disponível em: <https://partnershiponai.org/about/>. Acesso em: 07 maio 2023.

PRATES, B. P.; CHIARINI, T. Dependência na Era Digital: a divisão centro-periferia em Inteligência Artificial. **Liinc em Revista**, Rio de Janeiro, v. 20, n. 2, 2024. Disponível em: <https://revista.ibict.br/liinc/article/view/7305/7052>. Acesso em: 23 set. 2025.

ROBBINS, S. A misdirected principle with a catch: explicability for AI. **Minds and Machines**, [s.l.], v. 29, [s.n.], p. 495–514, 2019. Disponível em: <https://link.springer.com/article/10.1007/s11023-019-09509-3>. Acesso em: 19 mar. 2023.

SCHNEIDER, M. Ética, política e epistemologia: interfaces da informação. *In*: ALBAGLI, Sarita (org.). **Fronteiras da Ciência da Informação**, Brasília (DF): IBICT, 2013.

SIAU, K.; WANG, W. Artificial Intelligence (AI) Ethics: Ethics of AI and Ethical AI. **Journal of Database Management**, [s.l.], v. 31, n. 02, p. 74-87, mar. 2020. Disponível em: https://www.researchgate.net/publication/340115931_Artificial_Intelligence_AI_Ethics_Ethics_of_AI_and_Ethical_AI. Acesso em: 19 mar. 2023.

STAHL, B. C.; WRIGHT, D. Ethics and privacy in AI and Big Data: implementing responsible research and innovation. **IEEE Security & Privacy**, [s.l.], v. 15, n. 03, p. 22-33, maio/jun. 2018. DOI [10.1109/MSP.2018.2701164](https://doi.org/10.1109/MSP.2018.2701164).

STRYKER, C.; SCAPICCHIO, M. **O que é IA generativa?** [s.l.]: IBM, 2024. Disponível em: <https://www.ibm.com/br-pt/think/topics/generative-ai>. Acesso em: 28 fev. 2025.

TRAN, B. X. *et al.* Global evolution of research in Artificial Intelligence in health and medicine: a bibliometric study. **Journal of Clinical Medicine**, [s.l.], v. 8, n. 3, mar. 2019. DOI [10.3390/jcm8030360](https://doi.org/10.3390/jcm8030360).

VANTI, N. A. P. Da Bibliometria à Webometria: uma exploração conceitual dos mecanismos utilizados para medir o registro da informação e a difusão do conhecimento. **Ciência da Informação**, Brasília (DF), v. 31, n. 2, p. 152-162, maio/ago. 2002. Disponível em: <https://revista.ibict.br/ciinf/article/view/970/1007>. Acesso em: 12 fev. 2025.

VISSER, M.; VAN ECK, N. J.; WALTMAN, L. Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. **Quantitative Science Studies**, [s.l.], v.2, [s.n.], p.20–41, 2021. DOI https://doi.org/10.1162/qss_a_00112.

WHITTLESTONE, J. *et al.* The Role and Limits of Principles in AI Ethics: Towards a Focus on Tensions. *In*: AIES '19: PROCEEDINGS OF THE 2019 AAAI/ACM CONFERENCE ON AI, ETHICS, AND SOCIETY, 19., 2019, Honolulu. **Proceedings...** Honolulu: AIES, 2019. p. 195–200. DOI <https://doi.org/10.1145/3306618.3314289>.

WINNER, L. Do artifacts have politics? **Daedalus**, v. 109, n. 1, p. 121–136, 1980.

Yin, R. K. **Pesquisa qualitativa do início ao fim**. Porto Alegre: Penso, 2016.

ZHU, J.; LIU, W. A tale of two databases: the use of Web of Science and Scopus in academic papers. **Scientometrics**, [s.l.], v. 123, [s.n.], p. 321–335, Feb. 2020. Disponível em: A tale of two databases: the use of Web of Science and Scopus in academic papers. Acesso em: 05 maio 2023.

ZUBOFF, S. **A era do capitalismo de vigilância**. Rio de Janeiro: Editora Intrínseca, 2021.