Revista Ciência Aberta Lusófona
V.2, 2026
E-ISSN: 3051-7842

Comunicação

A Ciência Aberta na Era da Inteligência Artificial: Direções e Oportunidades de Pesquisa

Open Science in the Age of Artificial Intelligence: Directions and Research Opportunities

La Ciencia Abierta en la Era de la Inteligencia Artificial: Direcciones y Oportunidades de Investigación

Edi Carlos de Oliveira*
Universidade Estadual do Paraná (UNESPAR), Brasil
ORCID: https://orcid.org/0000-0002-6316-6142
Lattes: http://lattes.cnpq.br/3804359178223599
edi.oliveira@unespar.edu.br

Fabieli Borges
Universidade Estadual do Oeste do Paraná (UNIOESTE), Brasil
ORCID: https://orcid.org/0000-0002-0835-5311
Lattes: http://lattes.cnpq.br/5857244387101386
fabieli.borges@unioeste.br

Marcelo Roger Meneghatti
Universidade Estadual do Oeste do Paraná (UNIOESTE), Brasil
ORCID: https://orcid.org/0000-0002-5736-4230
Lattes: http://lattes.cnpq.br/3424071828229550
marcelo.meneghatti@unioeste.br

* Autor correspondente.

Resumo

O objetivo desta pesquisa é analisar como a interface entre Ciência Aberta e Inteligência Artificial (IA) tem se estruturado enquanto campo científico emergente. A amostra foi composta por 96 artigos indexados na Web of Science entre 2016 e 2025. Foram examinados os padrões de produtividade, dispersão editorial, diversidade conceitual e organização temática por meio de análise bibliométrica baseada nas Leis de Lotka (1926), Bradford (1934) e Zipf (1949) e no acoplamento bibliográfico de autores. Os resultados revelam um campo altamente fragmentado, forte dispersão editorial e diversidade terminológica, o que indica ausência de um núcleo consolidado de periódicos e conceitos. O acoplamento bibliográfico identificou oito *clusters* temáticos, com baixa interconexão estrutural na rede, distribuídos entre núcleos e nichos emergentes. Conclui-se que a interface entre Ciência Aberta e IA encontra-se em processo de consolidação, apresentando pluralidade conceitual e oportunidades de investigação futura, especialmente em temas como governança de dados, reprodutibilidade computacional e padronização metodológica, dimensões essenciais para que a IA gere conhecimento científico confiável, transparente e socialmente relevante.

Palavras-chave:

Ciência Aberta; Inteligência Artificial; Reprodutibilidade Científica; Governança de dados; Transparência científica; Consolidação científica.

Abstract

The objective of this research is to analyze how the interface between Open Science and Artificial Intelligence (AI) has been structured as an emerging scientific field. The sample consisted of 96 articles indexed in the *Web of Science* between 2016 and 2025. Patterns of productivity, editorial dispersion, conceptual diversity, and thematic organization were examined through bibliometric analysis based on Lotka's Law (1926), Bradford's Law (1934), and Zipf's Law (1949), and on the bibliographic coupling of authors. The results reveal a highly fragmented field, strong editorial dispersion, and terminological diversity, indicating the absence of a consolidated core of journals and concepts. Bibliographic coupling identified eight thematic *clusters*, with low structural interconnection in the network, distributed among emerging cores and niches. It is concluded that the interface between Open Science and AI is in the process of consolidation, presenting conceptual plurality and opportunities for future research, especially in areas such as data governance, computational reproducibility, and methodological standardization—essential dimensions for AI to generate reliable, transparent, and socially relevant scientific knowledge.

Keywords:

Open Science; Artificial Intelligence; Scientific Reproducibility; Data Governance; Scientific Transparency; Scientific Consolidation.

Resumen

El objetivo de esta investigación es analizar cómo se ha estructurado la interfaz entre la Ciencia Abierta y la Inteligencia Artificial (IA) como un campo científico emergente. La muestra consistió en 96 artículos indexados en la *Web of Science* entre 2016 y 2025. Se examinaron los patrones de productividad, dispersión editorial, diversidad conceptual y organización temática mediante un análisis bibliométrico basado en la Ley de Lotka (1926), la Ley de Bradford (1934) y la Ley de Zipf (1949), y en el acoplamiento bibliográfico de los autores. Los resultados revelan un campo altamente fragmentado, una fuerte dispersión editorial y diversidad terminológica, lo que indica la ausencia de un núcleo consolidado de revistas y conceptos. El acoplamiento bibliográfico identificó ocho clústeres temáticos, con baja interconexión estructural en la red, distribuidos entre núcleos y nichos emergentes. Se concluye que la interfaz entre Ciencia Abierta e IA está en proceso de consolidación, presentando pluralidad conceptual y oportunidades para futuras investigaciones, especialmente en áreas como gobernanza de datos, reproducibilidad computacional y estandarización metodológica, dimensiones esenciales para que la IA genere conocimiento científico confiable, transparente y socialmente relevante.

Palabras clave:

Ciencia Abierta; Inteligencia Artificial; Reproducibilidad Científica; Gobernanza de Datos; Transparencia Científica; Consolidación Científica.

1. Introdução

A Ciência Aberta tem assumido relevância global ao propor um modelo de produção e circulação do conhecimento que busca transparência, reprodutibilidade, colaboração e acesso aberto (Fecher & Friesike, 2014; UNESCO, 2023). Ao mesmo tempo, a Inteligência Artificial (IA) emerge como uma das tecnologias mais transformadoras da atualidade, influenciando processos produtivos e metodologias científicas, práticas de gestão de dados e dinâmicas de pesquisa (Floridi & Chiriatti, 2020; Russell & Norvig, 2021). A convergência entre ambas tem ampliado debates sobre ética, reprodutibilidade algorítmica, automação da pesquisa e infraestrutura de dados, configurando um campo emergente ainda pouco consolidado, embora em importante crescimento. Essa interface envolve desafios relacionados à transparência dos modelos, abertura

dos dados de treinamento, integridade dos processos computacionais e impacto das tecnologias inteligentes sobre práticas científicas.

Compreender como a relação entre Ciência Aberta e IA tem sido tratada na literatura é fundamental para identificar tendências, lacunas e oportunidades de pesquisa. Neste ínterim, a bibliometria é uma ferramenta que pode (co)responder a esta procura, sendo ela adequada para mapear padrões de produtividade, dispersão, frequência terminológica e estrutura intelectual (Chueke & Amatucci, 2015). O objetivo desta pesquisa é analisar como a interface entre Ciência Aberta e IA tem se estruturado enquanto campo científico emergente.

Para isso, realizou-se uma pesquisa quantitativa e descritiva baseada em análise bibliométrica. O *corpus* foi composto por 96 artigos indexados na Web of Science, publicados entre 2016 e 2025. A investigação examinou padrões de produtividade, dispersão editorial, diversidade terminológica e organização temática, fundamentando-se nas Leis de Lotka (1926), Bradford (1934) e Zipf (1949) e na análise de acoplamento bibliográfico, com apoio do software VOSviewer.

2. Ciência Aberta: princípios e transformações

A Ciência Aberta representa um movimento global que abrange práticas como acesso aberto, dados abertos, ciência cidadã, licenças abertas e reprodutibilidade científica (Vicente-Sáez & Martínez, 2018). A UNESCO (2023) destaca que o avanço da Ciência Aberta depende de infraestrutura tecnológica robusta, políticas institucionais e cultura de compartilhamento. A Ciência Aberta envolve novos modelos de disseminação e uso de dados, com forte dependência de padrões técnicos, ferramentas computacionais e plataformas de armazenamento, o que aproxima esse campo de estudos à temática de IA (Fecher & Friesike, 2014).

A Ciência Aberta deve ser compreendida como um conjunto de práticas operacionais e como uma mudança paradigmática na própria lógica de produção, circulação e validação do conhecimento científico. A emergência de ecossistemas digitais orientados por dados e algoritmos torna evidente que a abertura de informações, métodos, códigos, resultados e infraestrutura constitui um ideal normativo e uma condição para que a ciência permaneça verificável, colaborativa e disponível à sociedade (Russell & Norvig, 2021). Afinal, em torno do seu conceito multifacetado, a Ciência Aberta também abrange uma democratização do conhecimento por meio da colaboração entre pesquisadores (Medeiros & Sobral, 2025). Assim, ao aproximar-se da IA, a Ciência Aberta expande seu alcance, revelando-se um eixo estruturante para garantir transparência, reprodutibilidade e confiança pública em um cenário em que a pesquisa se torna cada vez mais automatizada e mediada por tecnologias computacionais.

A IA vem sendo utilizada para automatizar a coleta, análise e interpretação de dados, auxiliando na detecção de padrões, simulação de experimentos e produção de conhecimento (Jordan &

Mitchell, 2015). Na ciência, a IA apoia práticas como mineração de textos, processamento de linguagem natural, predição de fenômenos e gestão de grandes bancos de dados. A literatura evidencia que a IA assume múltiplos papéis no ecossistema científico contemporâneo: como ferramenta de automatização de processos; como mecanismo para gestão e mineração de dados, inferência e predição; e como vetor de transformação da própria dinâmica de produção do conhecimento (Russel & Norvig, 2021). Ao mesmo tempo, emergem discussões éticas fundamentais, especialmente relacionadas à transparência de modelos, vieses algorítmicos e explicabilidade (Mittelstadt et al., 2016).

A convergência entre Ciência Aberta e IA torna-se estratégica, uma vez que políticas e práticas de abertura, envolvendo algoritmos, dados e fluxos de pesquisa são essenciais para promover reprodutibilidade computacional e garantir que a inovação tecnológica permaneça alinhada à integridade científica e ao interesse público (Fecher & Friesike, 2014; Vicente-Sáez & Martínez-Fuentes, 2018; UNESCO, 2023). O avanço da IA na ciência não pode ser dissociado das condições sociotécnicas que viabilizam sua utilização de forma responsável, verificável e sustentável.

Se, por um lado, a IA amplia a capacidade humana de analisar fenômenos complexos e acelerar descobertas, por outro, sua efetividade depende de ambientes em que os dados sejam acessíveis, os processos reprodutíveis e os modelos auditáveis. Este último, se soma preocupação sob prisma político e de custos, reflexões estas também necessárias em especial para cenários públicos (Medeiros & Sobral, 2025). Segundo a literatura, a IA somente alcançará seu potencial transformador quando integrada a uma cultura de abertura e colaboração científica (pressupostos centrais da Ciência Aberta) que permita automatizar a ciência, democratizar seus benefícios e fortalecer sua legitimidade perante a sociedade. Sem práticas de Ciência Aberta, a IA aplicada à pesquisa científica tende a operar como uma tecnologia superficial, limitando a verificabilidade, a reprodutibilidade e a confiança pública nos resultados.

3. Procedimentos Metodológicos

Esta pesquisa caracteriza-se como um estudo com procedimentos de análise bibliométrica de natureza quantitativa, desenvolvido a partir de dados secundários extraídos da base Web of Science Core Collection (Clarivate Analytics). A bibliometria constitui uma abordagem consolidada para mapear a estrutura intelectual de um campo científico, identificar padrões de produção, colaboração e disseminação do conhecimento, bem como compreender sua evolução temática e conceitual (Zupic & Čater, 2015; Aria & Cuccurullo, 2017; Donthu et al., 2021). O corpus de análise foi composto por artigos científicos publicados entre 2016 e 2025.

A estratégia de busca consistiu na utilização do termo "open science" como expressão inicial de busca na base de dados selecionada, seguida da aplicação sucessiva de filtros temáticos e de tipo

de documento disponibilizados pela própria base. A coleta foi realizada em 5 de dezembro de 2025. Inicialmente, a busca pelo termo "open science" identificou 8.510 artigos. Em seguida, aplicou-se o filtro temático relacionado à "artificial intelligence", reduzindo o conjunto para 359 registros. Posteriormente, restringiu-se a busca ao tipo de documento "artigo científico", resultando em uma amostra de 96 artigos. Na fase de triagem e elegibilidade, foram analisados os resumos e as palavras-chave dos 96 artigos, não sendo identificadas exclusões adicionais nem registros duplicados. Assim, a amostra final foi composta por 96 artigos científicos, provenientes de 25 países, dos quais 94 foram publicados em inglês e dois em chinês, sendo todos incluídos na análise bibliométrica.

Reconhece-se que as publicações indexadas na Web of Science não esgotam a totalidade da produção científica sobre o tema. No entanto, a adoção dessa base de dados se justifica por sua abrangência, padronização e confiabilidade dos registros, e pela sua relevância notória na temática estudada. Assim, o conjunto analisado constitui um recorte analítico consistente, adequado aos objetivos do estudo e passível de subsidiar análises posteriores de caráter mais amplo. A partir da compilação dos indicadores bibliométricos, buscou-se analisar a Ciência Aberta pela lente teórica da IA, no período de 2016 a 2025. Embora a concentração das publicações ocorra entre 2022 e 2025, optou-se por analisar todo o período de 2016 a 2025 para captar o surgimento, a evolução e a consolidação gradual da interface entre Ciência Aberta e Inteligência Artificial. A inclusão dos anos iniciais possibilita identificar o momento de emergência do tema e compreender sua trajetória de crescimento ao longo da década.

No que se refere às técnicas de análise, foram empregadas as leis bibliométricas clássicas de Lotka (1926), Bradford (1934) e Zipf (1949), com o objetivo de examinar, respectivamente, a dispersão dos periódicos, a produtividade dos autores e a frequência dos termos associados à temática investigada. Adicionalmente, utilizou-se o acoplamento bibliográfico como medida da estrutura intelectual do campo, permitindo identificar agrupamentos de artigos com bases teóricas comuns. Para a operacionalização das análises e a visualização das redes bibliométricas, recorreu-se ao software VOSviewer, amplamente empregado em estudos dessa natureza (Van Eck & Waltman, 2010).

A combinação das leis bibliométricas clássicas de Lotka (1926), Bradford (1934) e Zipf (1949) com a análise de acoplamento bibliográfico mostrou-se adequada para caracterizar simultaneamente os padrões de produtividade, dispersão, estrutura temática e organização intelectual do campo investigado. Essa integração de técnicas é amplamente empregada em pesquisas bibliométricas por possibilitar uma compreensão abrangente da evolução e da configuração estrutural de domínios científicos (Zupic & Čater, 2015; Donthu et al., 2021).

4. Resultados

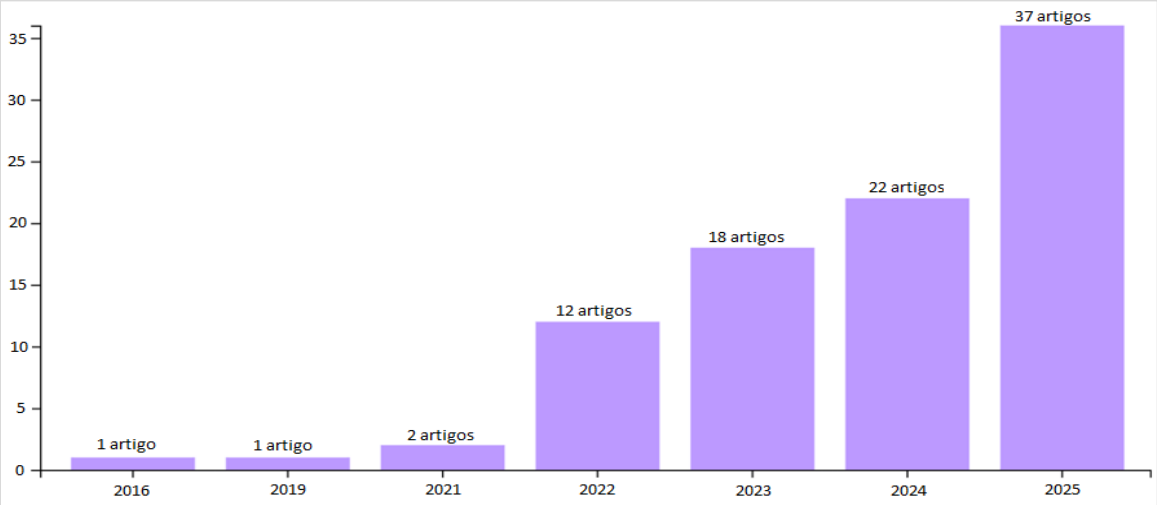
A análise da amostra permitiu identificar indicadores bibliométricos relevantes, tais como a distribuição temporal das publicações, evidenciando sua evolução ao longo do período analisado; os países responsáveis pela produção científica; os periódicos que concentraram os estudos; e os autores que contribuíram para essa literatura. Essas informações possibilitam uma visão panorâmica e contextualizada deste conjunto de publicações. A partir desse panorama inicial, as análises subsequentes aprofundam-se na aplicação das leis bibliométricas clássicas de Lotka (1926), Bradford (1934) e Zipf (1949), bem como na análise de acoplamento bibliográfico, com o objetivo de examinar padrões de produtividade, dispersão, recorrência temática e a estrutura intelectual que caracteriza o campo de pesquisa em Ciência Aberta na perspectiva da IA.

4.1. Lei de Lotka: produtividade dos autores

O Figura 1 apresenta a distribuição anual das publicações que compõem a amostra analisada, evidenciando a evolução do número de artigos publicados entre 2016 e 2025. A análise desse indicador temporal permite compreender a dinâmica de crescimento da produção científica relacionada à articulação entre Ciência Aberta e IA ao longo do período investigado.

Figura 1

Período de publicações dos 96 artigos da amostra (anos)



Nota: Web of Science (2025).

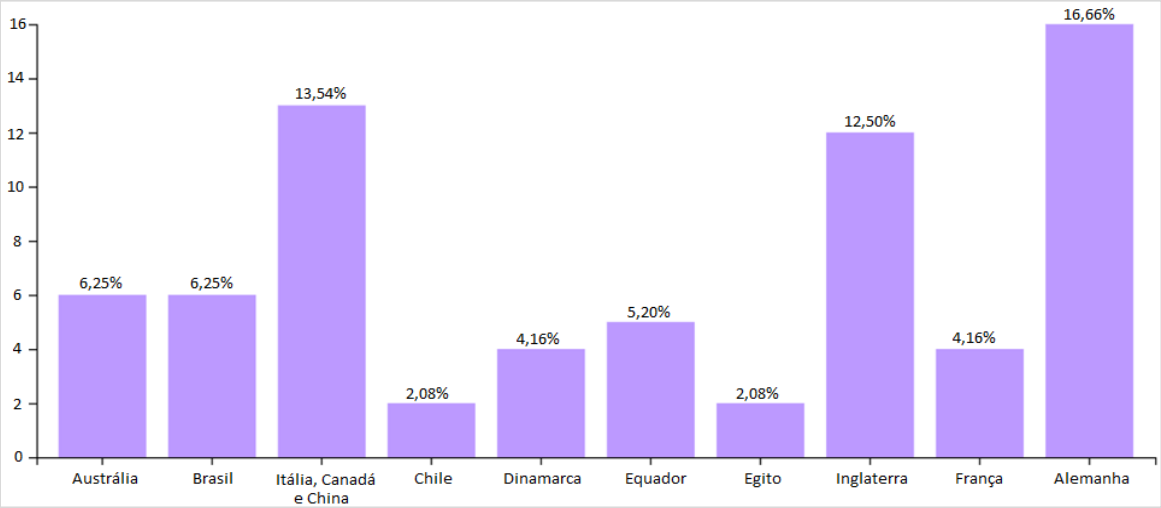
A distribuição temporal das publicações que compõem a amostra (Gráfico 1) evidencia um crescimento expressivo do interesse científico pelo tema ao longo do período analisado, 2016 a

2025. Observa-se que, entre 2016 e 2020, o número de artigos publicados foi incipiente, indicando uma fase inicial e ainda pouco explorada da literatura. A partir de 2021, nota-se uma inflexão no volume de publicações, com crescimento mais consistente a partir de 2022. Esse aumento torna-se mais acentuado nos anos subsequentes, com destaque para 2023 e 2024, culminando em 2025 como o ano com maior número de publicações da amostra. Tal comportamento sugere a consolidação e a intensificação das investigações que articulam Ciência Aberta e Inteligência Artificial, refletindo tanto a maturação do debate acadêmico quanto a ampliação do uso de tecnologias baseadas em IA nos processos científicos. O padrão observado indica, portanto, uma tendência de expansão recente e acelerada do campo, característica de áreas emergentes e em rápida evolução.

Sequencialmente, o Figura 2 apresenta a distribuição dos artigos da amostra segundo os países ou regiões de afiliação dos autores, considerando o total de 96 publicações analisadas, permitindo identificar a contribuição geográfica para o desenvolvimento das pesquisas que articulam Ciência Aberta e IA. Os 96 artigos são provenientes de apenas 25 países; e destes, 94 foram publicados em inglês e 2 foram publicados em chinês.

Figura 2

Países representados na amostra utilizada



Nota: Web of Science (2025).

A distribuição geográfica das publicações evidencia a predominância de países com sistemas científicos consolidados e forte investimento em pesquisa e inovação. Ressalta-se que, embora os Estados Unidos concentrem a maior parcela da produção científica analisada (33,33% dos artigos), esse país não foi representado no gráfico com o objetivo de evitar distorções visuais decorrentes

de sua elevada participação relativa. A exclusão dos Estados Unidos da representação gráfica permitiu uma melhor visualização e comparação entre os demais países, cujas contribuições, embora relevantes, apresentam valores proporcionalmente inferiores.

O Gráfico 2 destaca a produção científica da Alemanha (16,66%) e de Itália, Canadá e China, cada um com 13,54% das publicações. Observa-se também participação expressiva de países europeus, como a Inglaterra (12,50%), indicando forte impulsionamento do debate no contexto europeu. A presença de países latino-americanos, como Brasil (6,25%), Equador (5,20%) e Chile (2,08%), ainda que menos expressiva, sinaliza a difusão gradual da agenda de Ciência Aberta em contextos emergentes.

No que se refere à produtividade dos autores, a maioria destes apresentam apenas uma publicação, enquanto poucos concentram maior recorrência autoral. Esse comportamento, compatível com a Lei de Lotka (1926), indica núcleos produtivos restritos e ainda em formação, com liderança exercida por poucos autores e entrada progressiva de novos pesquisadores no campo. Tal configuração sugere que as pesquisas que articulam Ciência Aberta e IA se encontram em uma fase de estruturação, com liderança exercida por poucos autores, ao mesmo tempo em que novos pesquisadores passam a integrar o campo.

4.2 Lei de Bradford: dispersão dos periódicos

A análise da dispersão dos periódicos, à luz da Lei de Bradford (1934), revela que a produção científica sobre Ciência Aberta na Era da IA apresenta elevada pulverização. Os 96 artigos identificados foram distribuídos em 89 periódicos, dos quais 84 (94,3%) publicaram apenas um artigo, evidenciando a predominância de contribuições únicas. Apenas cinco periódicos concentraram mais de uma publicação: *Journal of Medical Internet Research* e *JMIR Research Protocols*, com três artigos cada, totalizando seis publicações (6,3%); e *Chinese Science Bulletin Chinese*, *Current Oncology* e *Drug Discovery Today*, com dois artigos cada (também totalizando seis publicações, com 6,3%).

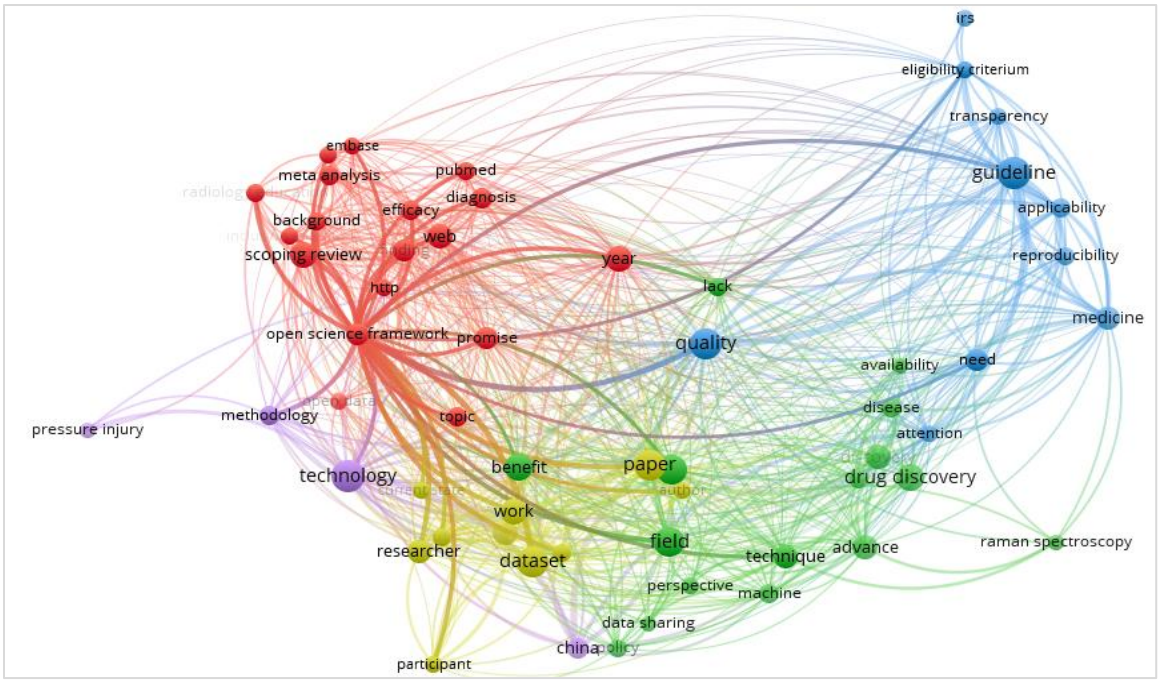
Essa configuração aponta que não há um núcleo consolidado de periódicos especializados no tema, sugerindo que a intersecção entre Ciência Aberta e IA ainda se encontra em processo de estruturação, distribuída entre áreas diversas e caracterizada por sua natureza interdisciplinar e emergente. A elevada dispersão observada confirma que a pesquisa em Ciência Aberta e IA ainda não consolidou periódicos de referência exclusivos, o que reforça a necessidade de fortalecimento institucional e de espaços editoriais dedicados ao tema.

4.3 Lei de Zipf: diversidade terminológica das palavras-chave dos artigos

A análise terminológica (Figura 3) mostra elevada variabilidade de termos, sugerindo amplo leque temático. Palavras relacionadas à saúde, biomedicina, ética e dados abertos aparecem com frequência, indicando convergência entre Ciência Aberta, IA e áreas biomédicas.

Figura 3

Diversidade terminológica das principais palavras-chave dos artigos da amostra



Nota: VOSviewer (2025).

A análise de coocorrência de 96 artigos revelou 1.976 palavras-chave. Destas, apenas 58 apresentaram frequência mínima de cinco ocorrências, formando o corpus do *VOSviewer*. Essa distribuição assimétrica confirma a Lei de Zipf (1949), onde poucos termos dominam semanticamente o campo científico.

A rede organizou-se em cinco *clusters* temáticos. O *Cluster* 1 (vermelho, 18 termos) foca na sistematização do conhecimento (ex: *meta analysis*), ocupando a centralidade da rede. O *Cluster* 2 (verde, 16 termos) abrange aplicações e dados (ex: *big data*, *drug discovery*), articulando metodologias e práticas.

Os *Clusters* 3 (azul, 10 termos) e 4 (amarelo, 10 termos) representam a frequência intermediária da distribuição de Zipf. O azul foca em diretrizes e reprodutibilidade (ex: *quality*, *transparency*), enquanto o amarelo detalha a estrutura operacional acadêmica (ex: *dataset*, *researcher*). O

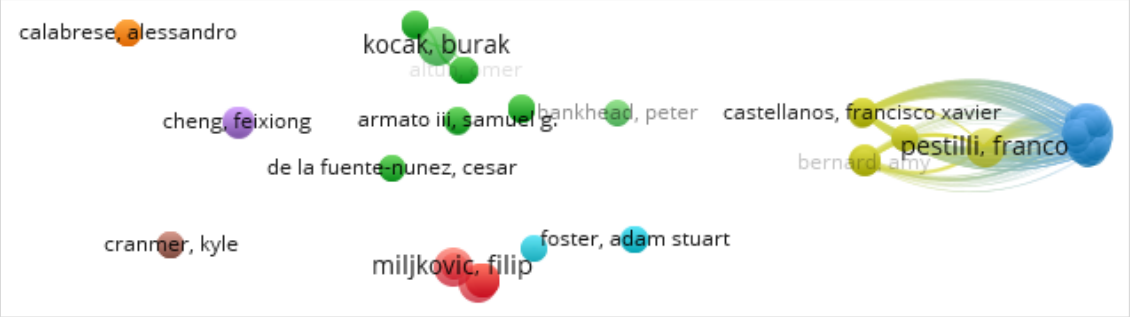
Cluster 5 (roxo, 4 termos), com itens como China e *pressure injury*, ilustra a "cauda longa" de Zipf (1949), com termos periféricos de baixa influência global. Essa hierarquia lexical encontrada nos *clusters*, fundamentada em Zipf (1949), prova que o domínio está consolidado em eixos centrais (sínteses e dados), mas mantém abertura para diversificação temática, característica comum em áreas científicas em expansão.

4.4. Acoplamento bibliográfico: *clusters* e fragmentação do campo

A análise realizada no *VOSviewer* identificou 125 autores distribuídos nos 96 artigos da amostra, entre os quais foi possível mapear relações de acoplamento bibliográfico. Considerando um ponto de corte mínimo de 10 citações compartilhadas, o acoplamento bibliográfico de autores resultou em oito *clusters* distintos (Figura 4). As conexões estabelecidas entre os itens da rede evidenciam o grau de proximidade temática entre pesquisadores, indicando como determinados autores convergem conceitualmente a partir das referências utilizadas em seus estudos.

Figura 4

Acoplamento bibliográfico da amostra



Nota: *VOSviewer* (2025).

A análise de acoplamento bibliográfico dos autores resultou em uma rede composta por oito *clusters* distintos, cada um representado por uma cor diferente. Essa fragmentação em múltiplos *clusters* indica que os autores da base de dados compartilham poucas referências em comum, formando pequenos grupos de afinidade teórica em vez de uma comunidade científica integrada. Observa-se que os *clusters* são formados por grupos reduzidos de autores. Isso significa que grande parte dos autores não compartilha referências suficientes para estabelecer ligações com outros membros da rede. Apenas alguns grupos apresentam conectividade mais evidente.

O *cluster* mais coeso visualmente é o grupo associado ao autor *Pestili, Franco*, que aparece conectado a outros autores próximos, formando um núcleo de colaboração teórica mais forte, com maior conectividade (*clusters* azul e amarelo). A presença de múltiplas ligações nesse

conjunto indica que esses autores utilizam bases bibliográficas semelhantes, sugerindo convergência temática ou alinhamento metodológico.

Outro agrupamento notável é o *cluster* central (verde) formado pelos autores *Kocak, Burak*; *Armato III, Samuel G.*; *Bankhead, Peter*; *De la Fuente Núñez, Cesar*, entre outros. Apesar de pequeno, esse grupo demonstra conexões claras, representando um segundo foco de compartilhamento de referências. Em contraste, alguns *clusters* são compostos por autores completamente isolados, como *Calabrese, Alessandro* (laranja) e *Cranmer, Kyle* (*cluster* castanho), que não apresentam acoplamentos com outros pesquisadores. Esses casos reforçam a dispersão temática da amostra e a baixa sobreposição nas bases bibliográficas dos estudos. Em síntese, a presença de oito *clusters* pequenos e desconectados revela que a produção analisada é altamente fragmentada, com poucas linhas de pesquisa interligadas. Apenas dois ou três *clusters* apresentam compartilhamento explícito de referências, enquanto o restante dos autores atua de maneira independente, do ponto de vista das citações utilizadas.

Embora a busca sistematizada na Web of Science tenha identificado 96 artigos, apenas um subconjunto desses estudos (20 artigos) foi efetivamente agrupado em *clusters* no *software VOSviewer* (6 artigos no *cluster* verde; 4 no vermelho; 4 no amarelo; 2 no azul-claro; 1 no azul; 1 no roxo; 1 no laranja; e 1 no castanho). Isso ocorre porque o *software* organiza redes com base em critérios mínimos de concorrência e força de ligação, incluindo apenas os itens que apresentam conexões significativas com outros estudos. Os demais artigos da amostra (76) embora relevantes, apresentaram baixa conectividade e permaneceram como nós periféricos ou isolados na rede completa.

A Ciência Aberta atua simultaneamente como suporte infraestrutural ao viabilizar acesso compartilhado a dados e recursos computacionais, e como motor de inovação metodológica, ao permitir validação, reprodução e difusão de algoritmos em larga escala. Ou seja, o *cluster* 8 indica que, embora ainda embrionárias, as técnicas de amostragem orientadas por IA têm potencial disruptivo não apenas para a física computacional, mas também para outras áreas científico-tecnológicas intensivas em dados, abrindo novas possibilidades de avanço na interface entre IA, métodos fundamentais e ecossistemas abertos de pesquisa. Em conjunto, os *clusters* revelam que a Ciência Aberta não opera como um paradigma unificador da IA científica, mas como um recurso instrumental mobilizado de forma contingente por diferentes comunidades disciplinares.

5. Discussão

A aplicação da Lei de Lotka (1926) confirma que o campo que articula Ciência Aberta e IA ainda se encontra em estágio de consolidação, marcado pela elevada proporção de autores com produção única e pela baixa concentração de pesquisadores recorrentes. Todavia, com produção

cronológica crescente, esse padrão, típico de áreas emergentes, indica a ausência de um núcleo epistemológico maduro capaz de sustentar continuidade teórica e agendas estáveis de pesquisa, o que contribui para a fragmentação temática observada. A recorrência limitada sugere uma entrada dispersa no campo, frequentemente associada a agendas específicas, como saúde, neurociência, farmacologia, governança de dados ou física computacional, sem integração consistente a uma comunidade científica unificada, tornando a formação de redes autorais densas um fator crítico para o avanço cumulativo e preferivelmente consumível do conhecimento.

Os resultados da Lei de Bradford (1934) reforçam esse diagnóstico ao evidenciar forte dispersão editorial: 96 artigos distribuídos em 89 periódicos, sendo que 94,3% publicaram apenas um estudo. A ausência de periódicos centrais confirma o caráter interdisciplinar da interface entre IA e Ciência Aberta e indica que o debate se desenvolve sobretudo em domínios onde a automação e a infraestrutura de dados são mais relevantes, como saúde, biomedicina, inovação farmacêutica e governança científica. Embora essa pulverização limite a consolidação de repertórios conceituais compartilhados, ela também amplia o alcance temático do campo, posicionando a Ciência Aberta como uma agenda transversal, capaz de atravessar diferentes territórios epistemológicos.

A análise da Lei de Zipf (1949) revela um domínio discursivo altamente heterogêneo, com 1.976 palavras-chave, das quais apenas 58 apresentam frequência significativa. Esse padrão evidencia a coexistência de eixos conceituais centrais, como dados abertos, *big data*, reprodutibilidade, Inteligência Artificial aplicada à saúde e descoberta de fármacos, e uma longa cauda de termos periféricos associados a nichos emergentes. A diversidade terminológica reflete tanto a experimentação metodológica quanto a ausência de estabilização conceitual, dificultando sínteses teóricas e comparabilidade de resultados, mas também indicando um ambiente fértil para inovação e incorporação de novas agendas conforme o campo amadurece. A pluralidade encontrada dentro das temáticas específicas não é apenas um estágio inicial, mas pode se tornar um risco estrutural se não houver, padrões, políticas e governança (UNESCO, 2023).

A distribuição dos oito *clusters* bibliográficos revela um campo plural e em expansão, composto por grandes núcleos teóricos, subgrupos consolidados e nichos especializados. De forma integrada, os *clusters* mostram que a Ciência Aberta, na era da IA, assume funções complementares que variam do papel de infraestrutura de dados e colaboração, à função normativa, ética e organizacional, até contextos em que sua adoção limitada restringe o avanço científico. Essa configuração evidencia que o fortalecimento da Ciência Aberta é condição estruturante para a legitimidade, a robustez metodológica e a sustentabilidade da IA potencializando a geração de conhecimento confiável, socialmente relevante e orientado por dados e algoritmos.

Do ponto de vista teórico, este estudo contribui ao demonstrar que a interface entre Ciência Aberta e IA ainda carece de um núcleo epistemológico consolidado, operando predominantemente como um campo de aplicações fragmentadas. Metodologicamente, o artigo avança ao integrar leis bibliométricas clássicas com análise de acoplamento bibliográfico, oferecendo uma leitura estrutural do campo. No plano prático e institucional, os achados reforçam a centralidade da governança de dados e da abertura metodológica como condições para a legitimidade científica da IA.

6. Considerações Finais

Este estudo teve como objetivo mapear a produção científica sobre a interface entre Ciência Aberta e Inteligência Artificial por meio de uma análise bibliométrica de artigos indexados na Web of Science. Os resultados evidenciam que esse campo ainda se encontra em consolidação, caracterizado por dispersão editorial, baixa recorrência de autores, elevada diversidade terminológica e reduzida integração entre parte significativa das publicações, embora apresente crescimento expressivo nos últimos anos.

A análise realizada demonstra que a Ciência Aberta desempenha um papel transversal no desenvolvimento da Inteligência Artificial, especialmente em áreas intensivas em dados, ao favorecer o compartilhamento de informações, a transparência, a reprodutibilidade e a colaboração científica. Ao integrar leis bibliométricas clássicas com a análise de acoplamento bibliográfico, o estudo contribui para compreender a estrutura e a evolução desse campo emergente, oferecendo subsídios para o desenvolvimento de novas pesquisas.

Este estudo apresenta algumas limitações. A primeira refere-se ao uso exclusivo da Web of Science, o que pode ter excluído publicações indexadas em outras bases, como Scopus, Dimensions ou Lens. Além disso, a estratégia de busca utilizou apenas o termo “open science”, associado posteriormente aos filtros temáticos da base, podendo não recuperar estudos que empregam terminologias correlatas. Outra limitação decorre da própria estrutura da rede bibliométrica, na qual parcela significativa dos artigos apresentou baixa conectividade, restringindo a formação de clusters e indicando um campo científico ainda fragmentado.

Como agenda futura, recomenda-se investigar o estudo de modelos de governança e políticas institucionais que incentivem práticas de Ciência Aberta em ambientes de pesquisa orientados por IA; a análise das barreiras culturais, técnicas e econômicas ao compartilhamento de dados e algoritmos; o desenvolvimento de métricas e indicadores para avaliação da reprodutibilidade algorítmica; a investigação dos impactos éticos e sociais da IA em ecossistemas científicos abertos; e a integração entre comunidades disciplinares atualmente dispersas, favorecendo a consolidação de agendas interdisciplinares mais coesas. Ou seja, o fortalecimento de práticas de abertura,

transparência e colaboração tende a desempenhar papel estratégico na consolidação desse campo e na produção de conhecimento científico mais confiável e socialmente relevante.

Conflito de Interesses

Os autores declaram não haver conflitos de interesses.

Disponibilização dos Dados de Investigação

Os dados utilizados nesta pesquisa encontram-se depositados no repositório Deposita Dados (IBICT), disponíveis em: https://doi.org/10.5281/zenodo.21326398

CRediT – Contribuições dos Autores

Edi Carlos de Oliveira | Conceitualização, Escrita – redação original, Supervisão, Aquisição de dados.
Fabieli Borges | Curadoria de dados, Escrita – revisão e edição.
Marcelo Roger Meneghatti | Escrita – revisão e edição, Supervisão.

Referências

Aria, M., & Cuccurullo, C. (2017). *bibliometrix*: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics, 11*(4), 959–975. https://doi.org/10.1016/j.joi.2017.08.007

Bradford, S. C. (1934). Sources of information on specific subjects. *Engineering: An Illustrated Weekly Journal* (London), 137, 85-86. https://doi.org/10.1177/016555158501000407

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research, 133,* 285–296. https://doi.org/10.1016/j.jbusres.2021.04.070

Fecher, B., & Friesike, S. (2014). Open Science: One Term, Five Schools of Thought. In Bartling, S., & Friesike, S., *Opening Science: The Evolving Guide on How the Internet is Changing Research, Collaboration and Scholarly Publishing* (pp. 17-47). Springer. https://doi.org/10.1007/978-3-319-00026-8_2

Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines, 30*(4), 681-694. https://doi.org/10.1007/s11023-020-09548-1

Jordan, M. I., & Mitchell, T. M. (2015). *Machine learning: Trends, perspectives, and prospects*. *Science, 349*(6245), 255-260. https://doi.org/10.1126/science.aaa8415

Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, *16*, 317–323. https://www.jstor.org/stable/24529203

Medeiros, C. B., & Sobral, F. A. F. (2025). Ciência Aberta: criação, disseminação e democratização do conhecimento. *Cienc. Cult. 77*(1). http://dx.doi.org/10.48207/2317-6660.20250001

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, *3*(2), 1-21. https://doi.org/10.1177/2053951716679679

Russell, S. J., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4th Ed.). Pearson Education. https://api.pageplace.de/preview/DT0400.9781292401171_A41586057/preview-9781292401171_A41586057.pdf

UNESCO. (2023). *Recommendation on Open Science.* https://www.unesco.org/en/open-science/about

Van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics, 84*(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3

Vicente-Sáez, R., & Martínez-Fuentes, C. (2018). Open Science now: A systematic literature review for an integrated. *Journal of Business Research*, *88*, 428-436. https://doi.org/10.1016/j.jbusres.2017.12.043

Zipf, G. K. (1949). *Human behavior and the principle of least effort*: *An introduction to human ecology*. Addison-Wesley. https://pure.mpg.de/rest/items/item_2407822_4/component/file_2562959/content

Zupic, I., & Čater, T. (2015). Bibliometric methods in management and organization. Organizational *Research Methods, 18*(3), 429–472. https://doi.org/10.1177/1094428114562629