

---

# Evolução e Estrutura da Produção Científica em Inteligência Artificial com Participação de Instituições Africanas (2018–2025): análise bibliométrica de publicações em acesso aberto indexadas na Web of Science e Scopus

*Evolution and Structure of Scientific Output in Artificial Intelligence Involving African Institutions (2018–2025): a bibliometric analysis of open-access publications indexed in Web of Science and Scopus*

---

**Arieta Domingas Bartolomeu Filipe (1), Geraldo Vernijo Deixa (2),  
Rosalino Subtil Chicote (3), Verônica Angélica Freitas de Paula (4)**

(1) Universidade Licungo (UniLicungo), Moçambique, arietafilipe3@gmail.com

(2) gdeixa@unilicungo.ac.mz

(3) Universidade Rovuma (UniRovuma), Moçambique, rchicote@unirovuma.ac.mz

(4) Universidade Federal de Uberlândia (UFU), Brasil, veronica@ufu.br



## Resumo

Este artigo caracteriza a evolução, a estrutura social e conceitual e o estágio de desenvolvimento da pesquisa sobre inteligência artificial (IA) na África. Realizou-se uma análise bibliométrica, com seleção dos documentos relatada segundo o PRISMA 2020, de artigos e revisões, em acesso aberto, indexados na Web of Science e na Scopus (2018–2025). Após a remoção de duplicados e a exclusão dos registros de 2026, o corpus integra 621 documentos, analisados com o pacote bibliometrix, em R. A taxa anual de crescimento calculada para o corpus é de 102,99% e concentra 45,7% dos documentos em 2025; o ajuste logístico exploratório ( $R^2 = 0,999$ ) estima um ponto de inflexão em 2027,5, com incerteza elevada. A Lei de Bradford descreve de forma aproximada a dispersão das fontes, enquanto a Lei de Lotka, na formulação clássica, não é validada ( $\beta = 4,87$ ). A África do Sul reúne 43,3% dos documentos, e o excedente de menções país-documento indica colaboração internacional frequente. A estrutura conceitual divide-se em duas vertentes, técnico-clínica e educativa, sem temas motores. O corpus apresenta concentração geográfica e baixa recorrência autoral; a extrapolação de seus resultados para toda a produção científica africana requer cautela.

**Palavras-chave:** inteligência artificial; África; bibliometria; leis bibliométricas; maturidade científica

---

FILIFE, Arieta Domingas Bartolomeu; DEIXA, Geraldo Vernijo; CHICOTE, Rosalino Subtil; PAULA, Verônica Angélica Freitas de. Evolução e Estrutura da Produção Científica em Inteligência Artificial com Participação de Instituições Africanas (2018–2025): análise bibliométrica de publicações em acesso aberto indexadas na Web of Science e Scopus. *Brazilian Journal of Information Science: research trends*, vol.20, publicação continua, 2026, e026039. DOI: <https://doi.org/10.36311/1981-1640.2026.v20.e026039>.

## Abstract

This article characterizes the evolution, the social and conceptual structure, and the maturity stage of research on artificial intelligence (AI) in Africa. A bibliometric analysis, with document selection reported according to PRISMA 2020, covered open-access articles and reviews indexed in Web of Science and Scopus (2018–2025). After duplicate removal and the exclusion of 2026 records, the corpus comprises 621 documents, analyzed with the bibliometrix package in R. Output grows by 102.99% per year and 45.7% of the documents were published in 2025; a logistic fit ( $R^2 = 0.999$ ) places the field before its inflection point, estimated at 2027.5. Bradford's Law describes the dispersion of sources only approximately, whereas Lotka's Law in its classical form is not validated ( $\beta = 4.87$ ). South Africa accounts for 43.3% of the documents, and the excess of country-document mentions indicates frequent international collaboration. The conceptual structure splits into a technical-clinical strand and an educational strand, with no motor themes. The field is young, concentrated, and fragmented, and consolidating recurrent authors, balanced networks, and themes that link technical performance, governance, and implementation is a priority.

**Keywords:** Artificial intelligence; Africa; Bibliometrics; Bibliometric laws; Scientific maturity

## 1 Introdução

---

A Inteligência Artificial (IA) tem sido empregada em diversas atividades, com aplicações em setores como a saúde, a agricultura, a educação, a governança e a proteção ambiental. No continente africano, o interesse pela IA expandiu-se de forma expressiva, impulsionado pela dupla necessidade de responder a desafios estruturais de desenvolvimento, tais como a insegurança alimentar, as assimetrias educativas e as vulnerabilidades climáticas, e de posicionar a região na economia digital global (Mienye *et al.*, 2024). Para compreender a evolução e a estrutura deste campo científico emergente, os estudos bibliométricos, fundamentados na aplicação de métodos estatísticos à produção escrita (Pritchard, 1969), têm desempenhado um papel crucial. Investigações preliminares documentam o crescimento acelerado da literatura africana sobre IA, indicando a liderança de polos como a África do Sul, a Nigéria e o Egito, com forte concentração em aplicações voltadas para as finanças e a saúde pública (Kondo; Diwani, 2023).

Apesar do crescente interesse e do aumento do número de publicações, a expansão da IA na África ocorre sob condições de infraestrutura digital limitada, escassez de dados de qualidade e ausência de quadros regulatórios robustos (Alaran *et al.*, 2025; Gikunda, 2024). A literatura recente aponta preocupações éticas relacionadas com o enviesamento algorítmico e com o risco de domínio externo sobre os dados, o chamado colonialismo digital (Alaran *et al.*, 2025), o que exige

modelos de governança alinhados com as realidades culturais locais. Embora existam mapeamentos gerais baseados nas leis de Lotka (1926), Bradford (1934) e Zipf (1949) e em técnicas contemporâneas de coocorrência de palavras e acoplamento bibliográfico (Cobo *et al.*, 2011; Donthu *et al.*, 2021), a produção científica atual carece de uma síntese crítica que relacione essas métricas quantitativas aos desafios socioeconômicos reais do continente. Subsiste uma deficiência na literatura, em que os dados de produtividade por país não dialogam diretamente com as barreiras estruturais e com as oportunidades de desenvolvimento sustentável reportadas empiricamente.

Diante desta lacuna, este estudo propõe-se ocupar esse espaço de pesquisa por meio de uma análise bibliométrica, com o bibliometrix (Aria; Cuccurullo, 2017). A abordagem é apresentada pelos autores em três dimensões: conceitual, ao unificar as métricas de redes de citações (Garfield, 1972; Small, 1973) com a análise do conteúdo temático da IA em contextos em desenvolvimento; prático, ao apresentar indicadores bibliométricos que podem subsidiar investigações sobre barreiras regulatórias e técnicas no continente; e político, ao identificar temas que podem ser examinados em estudos futuros.

Para orientar empiricamente esta pesquisa, formulou-se a seguinte pergunta central: como se caracterizam a evolução, a estrutura social e conceitual e a fase de maturidade da produção científica sobre IA com afiliação institucional em países africanos, indexada nas bases Scopus e Web of Science entre 2018 e 2025? O objetivo geral é caracterizar a evolução, a estrutura social e conceitual e o estágio de maturidade dessa produção. Definiram-se cinco objetivos específicos: (1) mensurar o volume anual, a taxa de crescimento e o impacto, medido pelas citações, das publicações; (2) ajustar um modelo de crescimento logístico para estimar a capacidade de saturação (K), o ponto de inflexão e a qualidade do ajuste ( $R^2$ , RMSE, AIC e BIC); (3) verificar a conformidade do campo com as leis de Lotka (produtividade dos autores) e de Bradford (dispersão das fontes); (4) analisar os padrões de autoria e colaboração nos níveis micro (autores) e macro (países); e (5) descrever a estrutura conceitual e a dinâmica temática a partir das palavras-chave.

## 2 Fundamentação teórica

---

A fundamentação teórica organiza-se em duas partes complementares. A primeira destaca os conceitos de IA. A segunda apresenta os fundamentos teóricos da bibliometria e as leis clássicas que sustentam a análise quantitativa da produção científica.

### 2.1 Conceitos de inteligência artificial

---

A IA tem sido amplamente compreendida como um campo da Ciência da Computação voltado para o desenvolvimento de sistemas capazes de executar tarefas que normalmente requerem inteligência humana. Nesse sentido, McCarthy (2007), reconhecido como um dos pioneiros da área, define a IA como a ciência e a engenharia de criar máquinas inteligentes, especialmente programas computacionais inteligentes. Segundo este autor, o propósito da IA consiste em compreender os mecanismos da inteligência humana e reproduzi-los em sistemas artificiais capazes de aprender, resolver problemas e tomar decisões. Essa concepção contribuiu para consolidar a ideia de que a inteligência pode ser simulada por meio de processos computacionais.

Entretanto, essa compreensão centrada na reprodução da inteligência humana é ampliada por Russell e Norvig (2021), que argumentam que a IA não deve ser entendida apenas como uma tentativa de imitar o pensamento humano. Segundo estes autores, a IA corresponde ao estudo de agentes inteligentes capazes de perceber o ambiente e agir racionalmente para alcançar determinados objetivos. Enquanto McCarthy enfatiza a criação de máquinas inteligentes, Russell e Norvig deslocam o foco para a capacidade dos sistemas de agir de forma eficiente e adaptativa em contextos complexos. Essa perspectiva sugere que a IA pode ser avaliada não apenas pela sua semelhança com a inteligência humana, mas também pelos resultados que consegue produzir.

Nessa perspectiva, Kaplan e Haenlein (2019) acrescentam que a IA envolve a capacidade de interpretar dados externos, aprender com eles e utilizar esse conhecimento para alcançar objetivos específicos. Os autores destacam a aprendizagem automática (*machine learning*) como um elemento central da IA contemporânea. Tal entendimento aproxima-se parcialmente da visão de Russell e Norvig ao enfatizar a adaptação dos sistemas ao ambiente. Contudo, enquanto estes

---

FILIPE, Arieta Domingas Bartolomeu; DEIXA, Geraldo Vernijo; CHICOTE, Rosalino Subtil; PAULA, Verônica Angélica Freitas de. Evolução e Estrutura da Produção Científica em Inteligência Artificial com Participação de Instituições Africanas (2018–2025): análise bibliométrica de publicações em acesso aberto indexadas na Web of Science e Scopus. *Brazilian Journal of Information Science: research trends*, vol.20, publicação continua, 2026, e026039. DOI: <https://doi.org/10.36311/1981-1640.2026.v20.e026039>.

privilegiam a racionalidade dos agentes, Kaplan e Haenlein enfatizam os processos de aprendizagem que permitem aos sistemas aperfeiçoar continuamente o seu desempenho a partir dos dados disponíveis.

Por outro lado, Luckin *et al.* (2016) propõem uma compreensão menos centrada na substituição das capacidades humanas e mais orientada para a sua ampliação. Os autores defendem que a IA deve ser entendida como um conjunto de sistemas computacionais capazes de complementar e ampliar capacidades cognitivas humanas, sobretudo em contextos de aprendizagem e de tomada de decisão. Essa posição contrasta com visões mais tradicionais, que associavam a IA à reprodução integral da inteligência humana. De modo semelhante, Nilsson (2010) argumenta que a IA consiste na atividade de tornar máquinas inteligentes, sendo a inteligência compreendida como a capacidade de agir adequadamente e com previsão em determinado ambiente. Ao enfatizar a autonomia e a adaptação dos sistemas inteligentes, o autor reforça a compreensão da IA como uma tecnologia dinâmica, capaz de responder às exigências de contextos variados.

Em síntese, embora os autores apresentem enfoques distintos, verifica-se uma convergência em torno da ideia de que a IA envolve a criação de sistemas capazes de aprender, raciocinar, adaptar-se e tomar decisões em contextos específicos. Assim, compreende-se que a IA evoluiu de uma concepção focada na simulação da inteligência humana para uma perspectiva mais ampla, que a reconhece como um conjunto de tecnologias capazes de apoiar, complementar e expandir as capacidades cognitivas humanas em diferentes domínios sociais, científicos e educacionais.

## 2.2 Fundamentos teóricos da bibliometria e leis clássicas

---

A denominação bibliometria foi proposta por Pritchard (1969), que a definiu como a aplicação de métodos matemáticos e estatísticos à produção científica, com o objetivo de descrever, avaliar e monitorar padrões de publicação, autoria e disseminação do conhecimento. Essa definição contribuiu para delimitar a bibliometria como um campo de investigação, distinto da bibliografia estatística que a precedeu, e abriu caminho ao desenvolvimento de um conjunto de

---

FILIPE, Arieta Domingas Bartolomeu; DEIXA, Geraldo Vernijo; CHICOTE, Rosalino Subtil; PAULA, Verônica Angélica Freitas de. Evolução e Estrutura da Produção Científica em Inteligência Artificial com Participação de Instituições Africanas (2018–2025): análise bibliométrica de publicações em acesso aberto indexadas na Web of Science e Scopus. *Brazilian Journal of Information Science: research trends*, vol.20, publicação continua, 2026, e026039. DOI: <https://doi.org/10.36311/1981-1640.2026.v20.e026039>.

leis empíricas que, décadas depois, continuam a orientar a caracterização quantitativa de campos científicos.

Entre estas leis clássicas destacam-se três contribuições fundamentais. A Lei de Lotka (Lotka, 1926) descreve a distribuição da produtividade científica entre autores, evidenciando que um pequeno número de pesquisadores é responsável por uma parcela desproporcionalmente elevada da produção de um campo, enquanto a maioria publica de forma esporádica. A Lei de Bradford (Bradford, 1934) descreve a dispersão da literatura científica entre as fontes, mostrando que um núcleo reduzido de revistas concentra a maior parte das publicações relevantes de uma área, o que permite identificar as fontes mais produtivas e orientar estratégias de pesquisa e de indexação. Por fim, a Lei de Zipf (Zipf, 1949) descreve a distribuição de frequência de termos em textos, fornecendo uma base para a análise de palavras-chave e a identificação de temas emergentes.

A estas leis juntou-se, a partir de meados do século XX, a tradição da análise de citações, iniciada por Garfield (1955) com a proposta dos índices de citação como instrumento para mapear as relações entre ideias na literatura científica. Essa ideia foi consolidada por Garfield (1972) com o desenvolvimento de indicadores bibliométricos para a avaliação de revistas e do impacto científico, entre os quais o conceito de fator de impacto.

A análise de citações evoluiu para o mapeamento científico (*science mapping*). Small (1973) introduziu o conceito de cocitação, que identifica relações temáticas entre documentos citados em conjunto. Callon *et al.* (1991) desenvolveram a análise de coocorrência de palavras como instrumento para descrever redes de interação entre pesquisa básica e tecnológica, e Cobo *et al.* (2011) propuseram uma abordagem para detectar, quantificar e visualizar a evolução temática de um campo de pesquisa.

Mais recentemente, a disponibilização de ferramentas computacionais acessíveis e de diretrizes metodológicas sistematizadas ampliou as possibilidades de aplicação da bibliometria a novas áreas do conhecimento. O pacote *bibliometrix*, desenvolvido para o ambiente R por Aria e Cuccurullo (2017), tornou-se uma ferramenta para a realização de análises de mapeamento

---

FILIPE, Arieta Domingas Bartolomeu; DEIXA, Geraldo Vernijo; CHICOTE, Rosalino Subtil; PAULA, Verônica Angélica Freitas de. Evolução e Estrutura da Produção Científica em Inteligência Artificial com Participação de Instituições Africanas (2018–2025): análise bibliométrica de publicações em acesso aberto indexadas na Web of Science e Scopus. *Brazilian Journal of Information Science: research trends*, vol.20, publicação continua, 2026, e026039. DOI: <https://doi.org/10.36311/1981-1640.2026.v20.e026039>.

científico de forma reprodutível. Paralelamente, Donthu *et al.* (2021) sistematizaram um conjunto de orientações práticas para a condução de análises bibliométricas, e Zupic e Čater (2015) publicaram um guia de referência sobre os métodos bibliométricos aplicados à gestão e às ciências sociais. Por fim, a articulação entre bibliometria e revisão sistemática da literatura, proposta por Tranfield *et al.* (2003) e sistematizada nas diretrizes PRISMA 2020 (Page *et al.*, 2021), estabeleceu um padrão metodológico que combina o rigor da seleção sistemática de documentos com a capacidade descritiva e analítica dos indicadores bibliométricos.

### 3 Métodos

---

Adotou-se uma análise bibliométrica quantitativa, precedida de identificação, triagem e seleção sistematizadas dos documentos. O relato do processo de seleção seguiu as diretrizes PRISMA 2020 (Page *et al.*, 2021), na linha proposta por Tranfield *et al.* (2003) para a construção de conhecimento com base em evidência sistematizada, e a análise bibliométrica seguiu as orientações de Donthu *et al.* (2021) e de Zupic e Čater (2015). O processamento dos dados foi realizado com o pacote bibliometrix (Aria; Cuccurullo, 2017), em ambiente R, incluindo sua interface Biblioshiny.

#### 3.1 Bases de dados e estratégia de pesquisa

---

A pesquisa foi realizada em 2026 nas bases Web of Science e Scopus, reconhecidas pela abrangência e pelo rigor na indexação de literatura multidisciplinar e cuja utilização combinada é recomendada por Donthu *et al.* (2021) para reduzir vieses de cobertura. A estratégia combinou os descritores artificial intelligence e machine learning com o termo Africa e com os nomes dos 54 países africanos, e foi aplicada ao campo de resumo (abstract) de cada base, para o período de 2018 a 2026. A string de busca completa consta do Apêndice A. A escolha dos dois descritores foi deliberada: procurou captar o núcleo conceitual mais abrangente do campo, evitando favorecer subáreas específicas da IA em detrimento da análise global da produção científica africana.

### 3.2 Critérios de elegibilidade e seleção dos documentos

Foram incluídos artigos originais e artigos de revisão, descritos pelos autores como revisados por pares, em acesso aberto, publicados entre 2018 e 2025, com afiliação institucional em países africanos. O Quadro 1 sintetiza o fluxo de seleção, adaptado do PRISMA 2020. A pesquisa identificou 5.993 registros na Web of Science e 9.526 na Scopus, em um total de 15.519. A aplicação dos critérios de inclusão nas próprias bases reduziu o conjunto para 773 registros. Após a importação e a integração das duas bases, com a função *mergeDbSources* do bibliometrix, foram identificados e removidos cinco duplicados, o que resultou em 768 registros únicos, triados por título e resumo.

Quadro 1 - Fluxo de seleção dos registros

| Fase              | Procedimento                                                                                        | Registros      |
|-------------------|-----------------------------------------------------------------------------------------------------|----------------|
| Identificação     | Pesquisa na <i>Web of Science</i>                                                                   | 5.993          |
| Identificação     | Pesquisa na <i>Scopus</i>                                                                           | 9.526          |
| Seleção nas bases | Aplicação dos critérios de inclusão (período, tipo de documento, acesso aberto, idioma e afiliação) | 773            |
| Integração        | Remoção de duplicados ( <i>mergeDbSources</i> )                                                     | -5; restam 768 |
| Triagem           | Exclusão dos registros de 2026, fora do recorte temporal                                            | -147           |
| Inclusão          | <i>Corpus</i> final (476 artigos e 145 revisões)                                                    | 621            |

Fonte: Adaptado de Page *et al.* (2021).

Dos 768 registros triados, foram excluídos 147 referentes a 2026, por estarem fora do recorte temporal definido (2018–2025), de modo a manter a série histórica fechada e a evitar o viés de um ano incompleto. A exclusão corresponde a 19,1% dos registros únicos triados e decorre de um critério temporal pré-definido, e não de uma seleção temática. Segundo o procedimento relatado pelos autores, a triagem por título e resumo não conduziu a outras exclusões, pelo que o *corpus* final reúne 621 documentos. O recorte até 2025 permite incluir o período posterior a 2022, marcado pela disseminação dos modelos de linguagem de grande porte (*large language models*), que permanece pouco explorado, já que os estudos bibliométricos anteriores sobre IA na África cobrem até 2022 (Kondo; Diwani, 2023).

### 3.3 Procedimentos de análise de dados

Os dados foram analisados com o *bibliometrix* e sua interface Biblioshiny, com cálculos complementares de indicadores derivados e de testes de aderência. O Quadro 2 relaciona cada eixo de resultados com os indicadores e os procedimentos aplicados.

Quadro 2 - Correspondência entre eixos de resultados, indicadores e procedimentos

| Eixo (seção)                       | Indicadores                                                                                                                                                | Procedimento                                                                                                                                                                                 |
|------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Produção, evolução e impacto (4.1) | Produção anual; taxa de crescimento anual; idade média; citações por documento e por ano; média de citações por ano de publicação; documentos mais citados | Estatística descritiva no <i>bibliometrix</i>                                                                                                                                                |
| Ciclo de vida (4.2)                | K, r e $t_0$ ; $R^2$ , RMSE, AIC e BIC; incremento máximo ( $K \cdot r/4$ ); intervalo entre 10% e 90% de K; projeções                                     | Ajuste não linear da curva logística aos oito totais anuais, com intervalos de confiança de 95%                                                                                              |
| Leis bibliométricas (4.3)          | Zonas de Bradford, multiplicador k e $R^2$ ; distribuição de Lotka, $\beta$ , C e $R^2$                                                                    | Bradford: três zonas com cerca de um terço dos documentos e ajuste da curva cumulativa. Lotka: regressão log-log e teste de Kolmogorov-Smirnov face a $\beta = 2$                            |
| Estrutura social (4.4)             | Autores por documento; documentos por autor; autoria única; documentos e citações por país                                                                 | Contagens a partir dos campos de autoria e afiliação; contagem total por país; citações por documento = citações ÷ documentos                                                                |
| Estrutura conceitual (4.5)         | Frequência de palavras-chave de autor (DE); rede de coocorrência (ID); mapa temático; dendrograma; tendência temporal (DE)                                 | Rede com identificação de <i>clusters</i> ; diagrama de centralidade e densidade; agrupamento hierárquico (método de Ward) das 30 palavras-chave indexadas mais frequentes; frequência anual |

Fonte: Elaborado pelos autores (2026)

Nos indicadores de produção e impacto, calcularam-se a produção anual, a taxa de crescimento anual composta, a idade média dos documentos (tomando 2026 como ano de referência), as citações médias por documento e por ano, a média de citações segundo o ano de publicação e a lista dos documentos mais citados. O impacto por país foi medido pelo total de

citações e pelo quociente entre citações e documentos. No ciclo de vida, a produção anual foi ajustada a uma curva logística,  $N(t) = K / (1 + e^{-(r(t - t_0))})$ , em que  $K$  é a capacidade de saturação,  $r$  a taxa de crescimento intrínseca e  $t_0$  o ponto de inflexão. A qualidade do ajuste foi avaliada por  $R^2$ , RMSE, AIC e BIC, e dos parâmetros derivaram-se o incremento anual máximo ( $K \cdot r/4$ ), o tempo de transição entre 10% e 90% de  $K$  ( $2 \cdot \ln 9 / r$ ) e as projeções anuais, interpretadas como cenários exploratórios, dada a dimensão da série.

Na Lei de Bradford, as fontes foram ordenadas por número de documentos e divididas em três zonas com cerca de um terço dos documentos cada; ajustou-se a curva cumulativa  $C(r) = a + b \cdot \ln(r)$ , da qual se obtiveram o  $R^2$  e o multiplicador  $k$ , e aplicou-se o teste de Kolmogorov-Smirnov. Na Lei de Lotka, a distribuição dos autores por número de documentos foi ajustada por regressão log-log,  $n(x) = C / x^\beta$ , e comparada com o expoente clássico ( $\beta = 2$ ) pelo teste de Kolmogorov-Smirnov, com valor crítico de  $1,63/\sqrt{N}$  ao nível de 1%. As frequências de autores com dois e três documentos foram obtidas por reconstrução, a partir dos totais de autores (3.553) e de aparições (3.877) e dos autores com quatro ou mais documentos.

Na estrutura social, calcularam-se, no nível micro, os autores por documento, os documentos por autor e a autoria única e, no nível macro, os documentos por país de afiliação, em contagem total (um documento conta para todos os países de afiliação), e as citações por país. Na estrutura conceitual, analisaram-se a frequência das palavras-chave de autor (DE) e sua evolução anual, a rede de coocorrência das palavras-chave indexadas (Keywords Plus, ID), o mapa temático, segundo o diagrama estratégico de centralidade e densidade (Callon *et al.*, 1991; Cobo *et al.*, 2011), e o agrupamento hierárquico das 30 palavras-chave indexadas mais frequentes pelo método de Ward. Os sinônimos não foram agregados. Por fim, os indicadores foram verificados por consistência interna, comparando totais anuais e globais e recalculando as medidas derivadas; os resultados que não puderam ser confirmados dessa forma não foram incluídos.

## 4 Resultados

Os resultados assentam em um corpus de 621 documentos sobre IA com afiliação institucional em países africanos, publicados de 2018 a 2025, obtidos pela fusão das bases Web of Science e Scopus e analisados com o pacote bibliometrix (Aria; Cuccurullo, 2017), em R. Organizam-se em cinco eixos: produção e impacto (4.1), ciclo de vida (4.2), leis bibliométricas (4.3), estrutura social (4.4) e estrutura conceitual (4.5).

### 4.1 Produção científica, evolução temporal e impacto

#### 4.1.1 Caracterização geral do corpus

A Tabela 1 resume as características do *corpus*. Os 621 documentos distribuem-se por 421 fontes e crescem a uma taxa anual de 102,99%, o que traduz a expansão muito rápida de um campo recente: a idade média dos documentos é de 2,09 anos (tomando 2026 como ano de referência) e 71,5% da produção concentra-se em 2024 e 2025.

Tabela 1 - Informações principais sobre o *corpus* (2018–2025)

| Indicador                                    | Valor     |
|----------------------------------------------|-----------|
| Período                                      | 2018–2025 |
| Fontes (revistas, livros etc.)               | 421       |
| Documentos                                   | 621       |
| Taxa de crescimento anual                    | 102,99%   |
| Idade média dos documentos (anos)            | 2,09      |
| Citações médias por documento                | 16,37     |
| Citações médias por documento e por ano      | 4,379     |
| Artigos                                      | 476       |
| Revisões                                     | 145       |
| Palavras-chave indexadas (Keywords Plus, ID) | 3.542     |
| Palavras-chave dos autores (DE)              | 2.161     |

Fonte: Dados da pesquisa (2026)

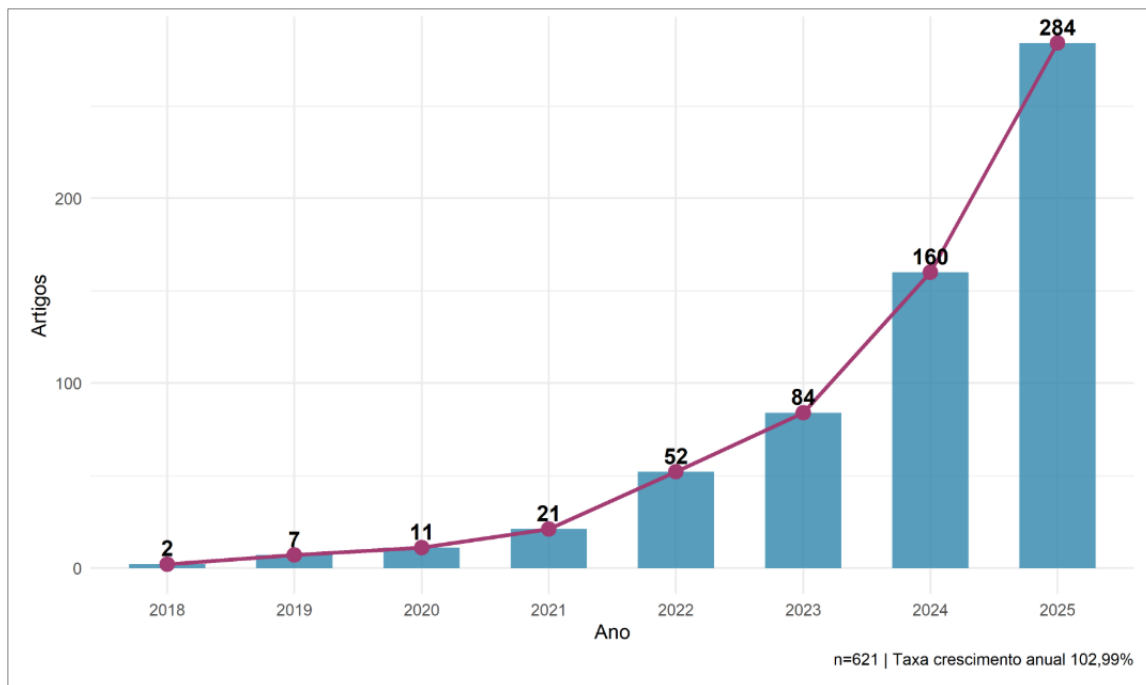
A tipologia é dominada por artigos originais (476), com 145 revisões, o que corresponde a uma razão de aproximadamente 3,3:1 entre artigos originais e revisões. O impacto médio é de

16,37 citações por documento (4,379 por documento e por ano). A base terminológica é ampla, com 3.542 palavras-chave indexadas e 2.161 palavras-chave de autor, o que sugere uma elevada diversidade temática. Os padrões de autoria e colaboração são analisados na Seção 4.4.

#### 4.1.2 Evolução temporal da produção

A produção anual (Figura 1) cresce lentamente entre 2018 e 2021, de 2 para 21 documentos por ano (41 documentos, 6,6% do corpus), e acelera a partir de 2022 (52 documentos). Em 2024 e 2025 foram publicados 160 e 284 documentos, respectivamente, e este último ano concentra sozinho 45,7% do corpus.

Figura 1 - Evolução da produção científica sobre IA na África (2018–2025)



Fonte: Dados da pesquisa (2026).

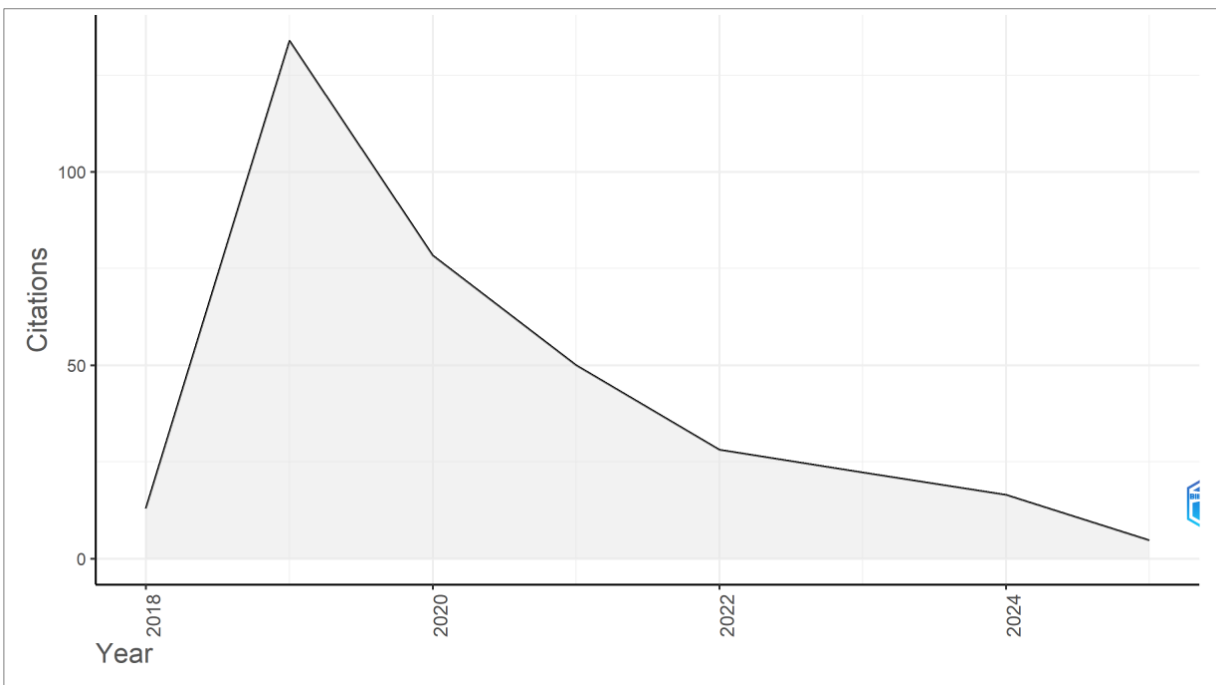
A Figura 1 evidencia que as variações anuais são sempre elevadas. O aumento de 250% em 2019 decorre de uma base muito baixa (2 documentos), mas as subidas de 147,6% em 2022 e de 90,5% em 2024 representam mudanças expressivas no volume de publicações, que marcam duas fases de aceleração. O incremento absoluto continua a aumentar (de +32 documentos em

2023 para +76 em 2024 e +124 em 2025), ainda que a variação relativa tenha diminuído (de 90,5% para 77,5%). O enquadramento desta trajetória em um modelo de ciclo de vida é feito na Seção 4.2.

#### 4.1.3 Impacto e documentos mais citados

O impacto, medido pela média de citações segundo o ano de publicação (Figura 2), é máximo nos documentos de 2019 (cerca de 134 citações por documento) e diminui de forma regular até 2025 (cerca de 5), passando por cerca de 78 em 2020, 50 em 2021, 28 em 2022 e 16 em 2024 (valores lidos na figura). Essa diferença entre anos de publicação é compatível com o tempo desigual de exposição a citações, já que os documentos recentes dispõem de menos tempo para acumular citações. O pico de 2019 é amplificado por dois documentos: os dois mais citados do *corpus*, ambos desse ano, somam 735 citações, o que equivale a 105 citações por documento quando repartidas pelos sete documentos publicados em 2019.

Figura 2 - Média de citações por ano de publicação



Fonte: Dados da pesquisa (2026).

Os dez documentos mais citados (Tabela 2) acumulam 2.519 citações e cobrem domínios muito distintos: agricultura, saúde, educação, recursos hídricos, finanças, turismo e gestão. Metade foi publicada em 2019–2020, mas dois trabalhos de 2024 (Abdaljaleel *et al.*, 2024; Booyse; Scheepers, 2024) já integram a lista, o que ilustra a rapidez da expansão posterior a 2022.

Tabela 2 - Dez documentos mais citados

| Pos. | Documento                           | Revista                                                                    | Domínio e âmbito                                                                            | Citações |
|------|-------------------------------------|----------------------------------------------------------------------------|---------------------------------------------------------------------------------------------|----------|
| 1    | Defourny <i>et al.</i> (2019)       | <i>Remote Sensing of Environment</i>                                       | Agricultura (sistema Sen2-Agri); locais de demonstração em vários países                    | 385      |
| 2    | Bellemo <i>et al.</i> (2019)        | <i>The Lancet Digital Health</i>                                           | Saúde (rastreamento de retinopatia diabética por deep learning); Zâmbia                     | 350      |
| 3    | Essel <i>et al.</i> (2022)          | <i>International Journal of Educational Technology in Higher Education</i> | Educação (chatbot como assistente de ensino); Gana                                          | 348      |
| 4    | Nti <i>et al.</i> (2020)            | <i>Journal of Big Data</i>                                                 | Finanças (previsão bolsista); dados de bolsas do Gana, dos EUA, da África do Sul e da Índia | 281      |
| 5    | Kouadri <i>et al.</i> (2021)        | <i>Applied Water Science</i>                                               | Recursos hídricos (índice de qualidade da água); Argélia                                    | 266      |
| 6    | Abdaljaleel <i>et al.</i> (2024)    | <i>Scientific Reports</i>                                                  | Educação (atitudes e uso do ChatGPT); estudo multinacional                                  | 227      |
| 7    | Gomez Selvaraj <i>et al.</i> (2020) | <i>ISPRS Journal of Photogrammetry and Remote Sensing</i>                  | Agricultura (detecção de doenças da bananeira); RDC e Benim                                 | 219      |
| 8    | Bhaga <i>et al.</i> (2020)          | <i>Remote Sensing</i>                                                      | Recursos hídricos e clima (revisão); África Subsaariana                                     | 157      |
| 9    | Al Fararni <i>et al.</i> (2021)     | <i>Big Data Mining and Analytics</i>                                       | Turismo (sistema de recomendação híbrido); Marrocos                                         | 148      |
| 10   | Booyse e Scheepers (2024)           | <i>Management Research Review</i>                                          | Gestão (barreiras à adoção de IA); África do Sul                                            | 138      |

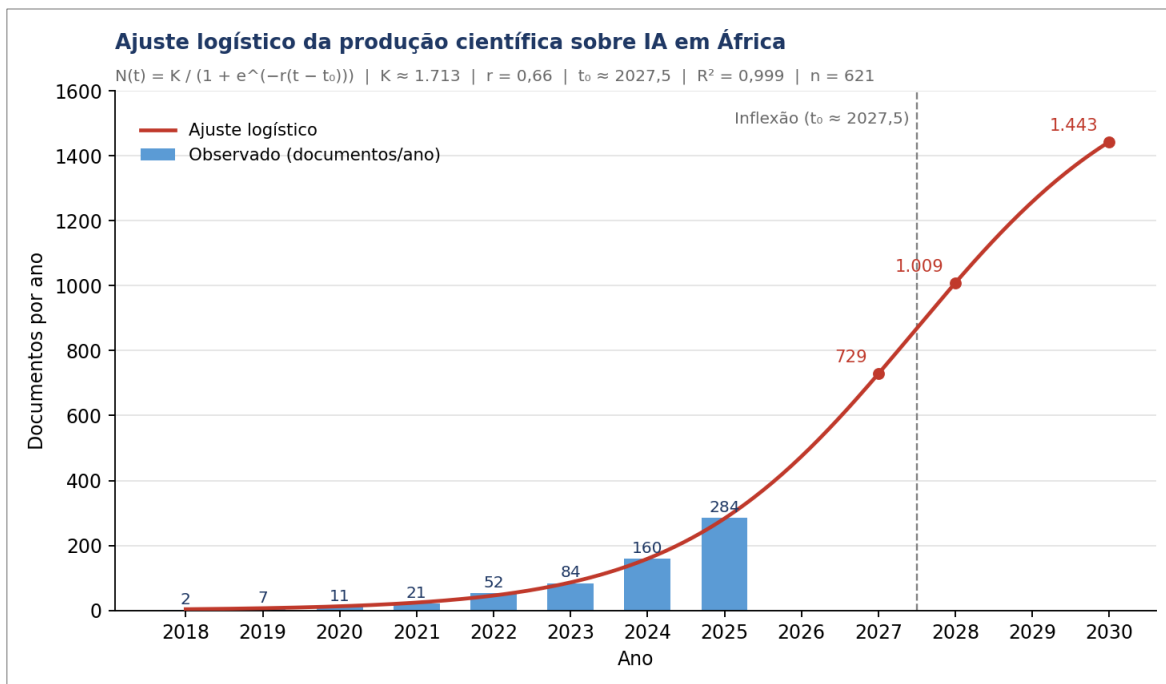
Fonte: Dados da pesquisa (2026).

Três dos dez estudos têm âmbito global ou multinacional (Defourny *et al.*, 2019; Nti *et al.*, 2020; Abdaljaleel *et al.*, 2024), o que indica que o corpus inclui trabalhos em que África é apenas um dos contextos de aplicação.

## 4.2 Ciclo de vida e estágio de maturidade

Para explorar a trajetória de crescimento, a produção anual foi ajustada a uma curva logística,  $N(t) = K / (1 + e^{-(r(t - t_0))})$ , em que  $K$  representa o limite assintótico estimado para a produção anual dentro do modelo,  $r$  a taxa de crescimento intrínseca e  $t_0$  o ponto de inflexão. O ajuste, realizado sobre os 621 documentos agregados em oito observações anuais, é muito elevado ( $R^2 = 0,999$ ;  $RMSE = 2,68$ ;  $AIC = 46,5$ ;  $BIC = 46,8$ ), e a curva reproduz de perto a trajetória observada (Figura 3).

Figura 3 - Ajuste logístico da produção científica sobre IA na África (2018–2025)



Fonte: Dados da pesquisa (2026).

A taxa de crescimento é  $r = 0,66$ , o ponto de inflexão situa-se em 2027,5 (IC 95%: 2024,8–2030,1) e o incremento anual máximo é de cerca de 282 publicações ( $K \cdot r/4$ ). No modelo ajustado, o limite assintótico é de aproximadamente 1.713 publicações por ano, mas com elevada incerteza: o limite inferior do intervalo de confiança é de cerca de 702 e não existe limite superior finito. A produção de 2025 (284 documentos) corresponde a cerca de 17% de  $K$ , e a passagem de 10% para

90% de K demora cerca de 6,7 anos. No âmbito do modelo ajustado, estes valores situam o período observado antes do ponto de inflexão; não comprovam que o campo efetivamente alcançará a saturação estimada. O modelo projeta 729 publicações em 2027, 1.009 em 2028 e 1.443 em 2030, ano em que se atingiriam 84% de K.

Dada a pequena dimensão da série (oito pontos para três parâmetros) e a incerteza de K, as projeções devem ser lidas como cenários exploratórios. A interpretação de crescimento recente também se apoia em indicadores descritivos: a idade média de 2,09 anos, os 45,7% da produção concentrados em 2025 e a ausência de temas motores no mapa temático (Seção 4.5). O  $R^2$  elevado, por si só, não valida a forma logística nem permite predizer uma saturação futura, sobretudo porque o ponto de inflexão estimado se situa fora do período observado.

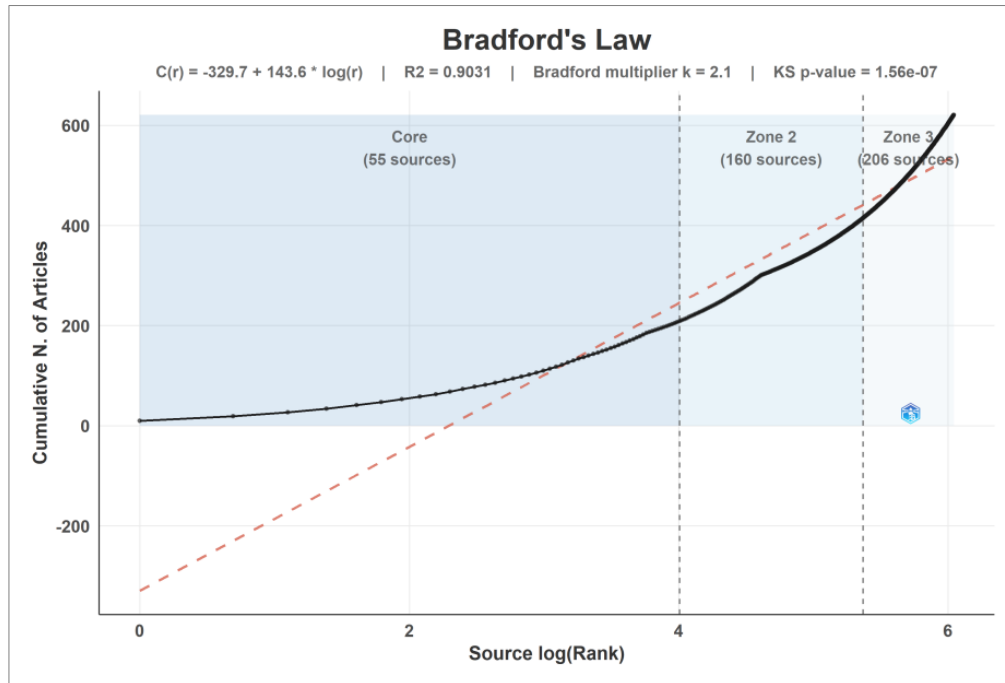
#### 4.3 Teste e validação das leis bibliométricas

---

##### 4.3.1 *Lei de Bradford: dispersão por fontes*

A dispersão dos 621 documentos pelas 421 fontes foi analisada com a Lei de Bradford (Figura 4). O método divide as fontes em três zonas que reúnem, cada uma, cerca de um terço dos documentos ( $\approx 207$ ): o núcleo é formado por 55 fontes (13,1% do total), a segunda zona por 160 (38,0%) e a terceira por 206 (48,9%).

Figura 4 - Zonas de Bradford (Lei de Bradford)



Fonte: Dados da pesquisa (2026).

O ajuste da curva cumulativa,  $C(r) = -329,7 + 143,6 \cdot \log(r)$ , tem  $R^2 = 0,903$ . O multiplicador de Bradford é  $k = 2,1$ , valor médio de dois quocientes entre zonas que não são constantes ( $160/55 = 2,9$  e  $206/160 = 1,3$ ), ao contrário do que a formulação clássica prevê. Em conjunto com o teste de Kolmogorov-Smirnov ( $p = 1,56 \times 10^{-7}$ ), que indica desvio estatisticamente significativo face ao modelo teórico, isto mostra que a lei descreve apenas de forma aproximada a dispersão observada. A concentração é, assim, real mas moderada. As dez fontes mais produtivas reúnem 68 documentos (10,95% do corpus): Scientific Reports (10), PLOS ONE (9), PLOS Neglected Tropical Diseases (8), Remote Sensing (7), International Journal of Learning, Teaching and Educational Research (7), Potchefstroom Electronic Law Journal (6), Discover Artificial Intelligence (6), Discover Sustainability (5), Cogent Education (5) e BMC Medical Education (5). A fonte mais produtiva, a Scientific Reports, tem apenas 1,6% dos documentos. O perfil das fontes é heterogêneo: revistas multidisciplinares de grande escala, de saúde e doenças tropicais, de detecção remota e sustentabilidade, de educação, de direito e de IA.

#### 4.3.2 Lei de Lotka: produtividade autoral

A produtividade dos 3.553 autores (3.877 aparições) foi analisada com a Lei de Lotka. A distribuição observada (Tabela 3) é muito concentrada nos autores ocasionais: 92,03% (3.270) assinam um só documento, 7,06% (251) dois e apenas 0,06% (2 autores) cinco ou mais. A lei clássica ( $\beta = 2$ ) previria 60,8% de autores com um só documento.

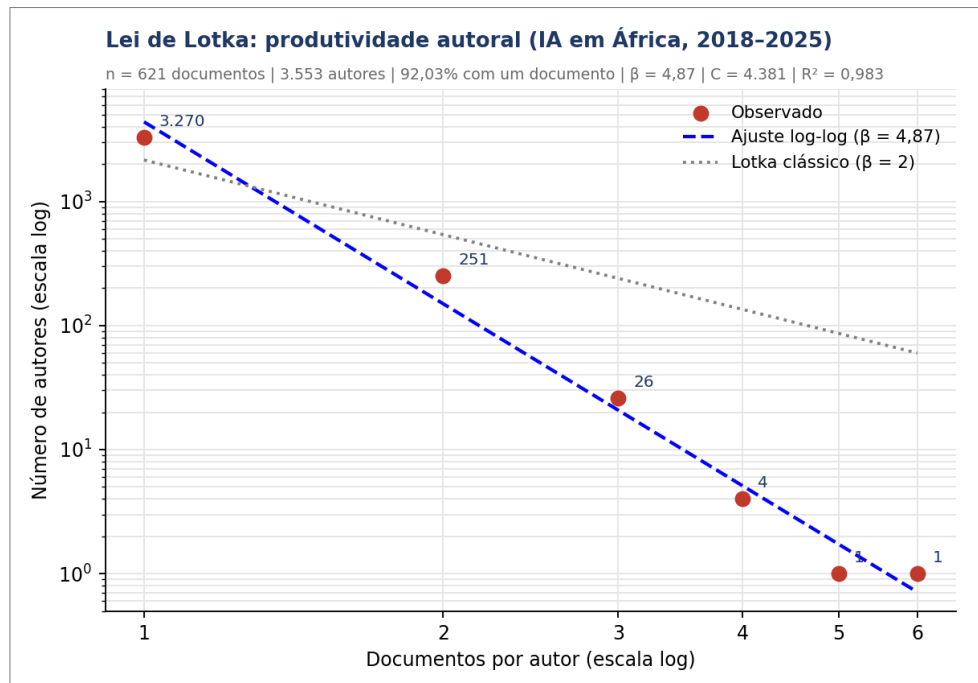
Tabela 3 - Distribuição da produtividade autoral e valores esperados pela Lei de Lotka clássica

| Documentos por autor | Autores (n)  | % observada   | % esperada ( $\beta = 2$ ) |
|----------------------|--------------|---------------|----------------------------|
| 1                    | 3.270        | 92,03         | 60,79                      |
| 2                    | 251          | 7,06          | 15,20                      |
| 3                    | 26           | 0,73          | 6,75                       |
| 4                    | 4            | 0,11          | 3,80                       |
| 5                    | 1            | 0,03          | 2,43                       |
| 6                    | 1            | 0,03          | 1,69                       |
| <b>Total</b>         | <b>3.553</b> | <b>100,00</b> | —                          |

Fonte: Dados da pesquisa (2026) (frequências para dois e três documentos reconstruídas a partir dos totais).

O ajuste log-log (Figura 5) resulta em  $\beta = 4,87$ ,  $C = 4.381$  e  $R^2 = 0,983$ , ou seja, em um decaimento muito mais acentuado do que o da lei clássica. O teste de Kolmogorov-Smirnov rejeita a aderência ao modelo clássico ( $D = 0,312$ , para um valor crítico de 0,027 a 1%). A distribuição segue, portanto, uma lei de potência, mas com um expoente muito superior a 2, compatível com uma comunidade formada sobretudo por autores ocasionais.

Figura 5 - Lei de Lotka: produtividade autorial em IA na África (2018–2025)



Fonte: Dados da pesquisa (2026).

A produção por autor apresenta baixa recorrência. O autor mais produtivo, Ayanwale MA, assina 6 documentos (0,97% do corpus), seguido de Chehbouni A (5) e de Dube T, Mabhaudhi T, Mijwil MM e Van Heerden A (4 cada). Em síntese, a Lei de Bradford descreve a dispersão das fontes de forma aproximada e a Lei de Lotka, na sua formulação clássica, não é validada.

#### 4.4 Estrutura social e redes de colaboração

A estrutura social da produção é analisada em dois níveis: o nível micro, relativo aos autores, e o nível macro, relativo aos países de afiliação.

##### 4.4.1 Nível micro: autores

Os 3.553 autores totalizam 3.877 aparições, o que equivale a 6,24 autores por documento e a 0,175 documentos por autor (Tabela 4). A autoria única é residual: 47 autores assinam os 48 documentos de autoria única (7,7% do corpus). A produção caracteriza-se, portanto, pela coautoria frequente e conduzida por equipes relativamente grandes.

Tabela 4 - Indicadores de autoria (nível micro)

| Indicador                              | Valor     |
|----------------------------------------|-----------|
| Autores                                | 3.553     |
| Aparições de autores                   | 3.877     |
| Autores por documento                  | 6,24      |
| Documentos por autor                   | 0,175     |
| Documentos de autoria única            | 48 (7,7%) |
| Autores de documentos de autoria única | 47        |

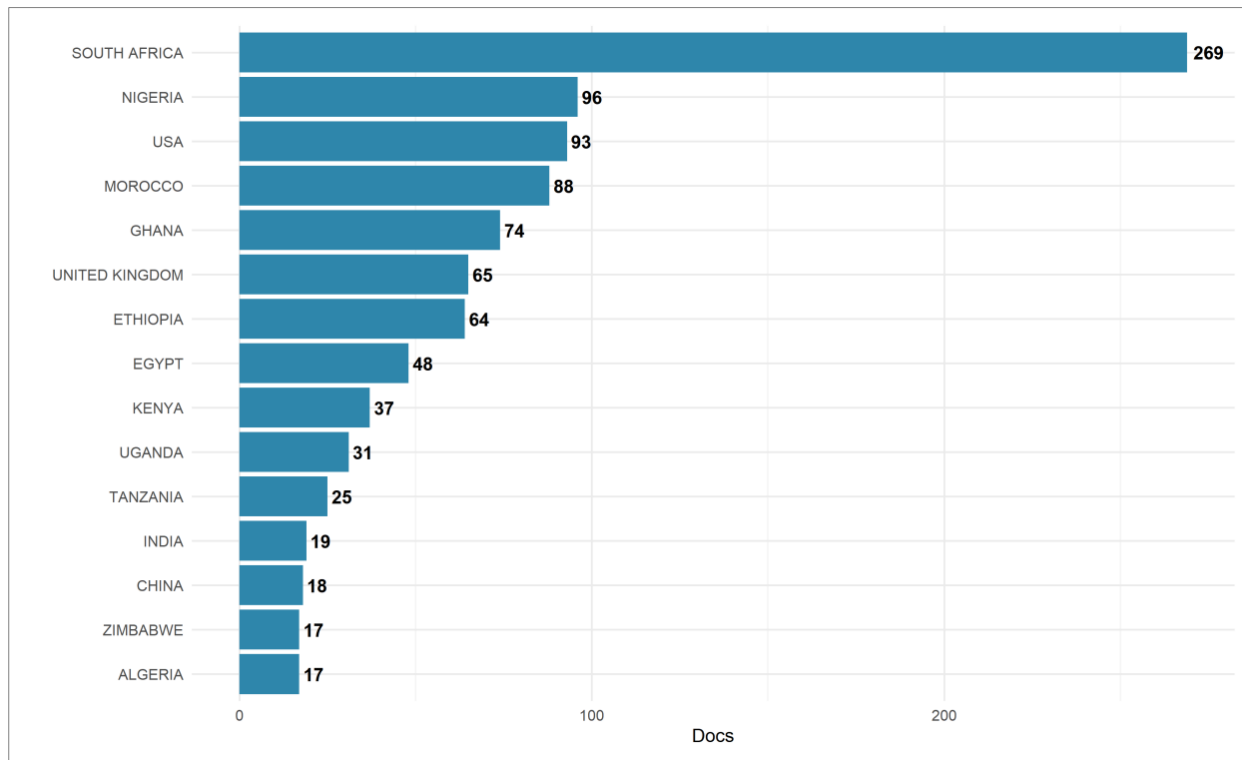
Fonte: Dados da pesquisa (2026).

Combinada com a elevada proporção de autores ocasionais (Seção 4.3.2), esta estrutura sugere equipes de composição variável e de baixa recorrência: poucos autores acumulam produção (o máximo é de seis documentos) e a maioria participa em um único estudo.

#### 4.4.2 Nível macro: países

A produção por país (Figura 6) é muito concentrada. A África do Sul lidera com 269 documentos (43,3% do corpus), cerca de 2,8 vezes o total da Nigéria, segunda colocada. Seguem-se a Nigéria (96; 15,5%), os EUA (93; 15,0%), Marrocos (88; 14,2%) e o Gana (74; 11,9%). Os EUA e o Reino Unido (65) figuram entre os seis primeiros, com 158 menções em conjunto, equivalentes a 25,4% do corpus.

Figura 6 - Distribuição da produção científica por país de afiliação (top 15)



Fonte: Dados da pesquisa (2026).

O segundo estrato inclui a Etiópia (64), o Egito (48), o Quênia (37), o Uganda (31) e a Tanzânia (25), e fecham o top 15 a Índia (19), a China (18), o Zimbábue (17) e a Argélia (17). A contagem é total, isto é, um documento conta para todos os países de afiliação: as 961 menções do top 15 excedem em 340 os 621 documentos, o que é compatível com coafiliação internacional, embora essa diferença, isoladamente, não permita estimar a proporção de documentos com colaboração internacional. O impacto por país (Tabela 5) revela uma dissociação entre volume e visibilidade. A África do Sul domina em citações absolutas (4.368; 31,4% do total dos dez países listados), mas as médias por documento são mais elevadas no Egito (26,1) e no Gana (22,9), e também na Itália que, ficando fora do top 15 de produção (no máximo 17 documentos), tem pelo menos 24,6 citações por documento. A Nigéria, com 96 documentos, registra apenas 6,6 citações por documento.

Tabela 5 - Dez países mais citados no *corpus*

| Pos. | País          | Total de citações | Documentos | Citações por documento |
|------|---------------|-------------------|------------|------------------------|
| 1    | África do Sul | 4.368             | 269        | 16,2                   |
| 2    | Gana          | 1.696             | 74         | 22,9                   |
| 3    | EUA           | 1.618             | 93         | 17,4                   |
| 4    | Marrocos      | 1.466             | 88         | 16,7                   |
| 5    | Egito         | 1.252             | 48         | 26,1                   |
| 6    | Reino Unido   | 1.081             | 65         | 16,6                   |
| 7    | Etiópia       | 914               | 64         | 14,3                   |
| 8    | Nigéria       | 634               | 96         | 6,6                    |
| 9    | Quênia        | 475               | 37         | 12,8                   |
| 10   | Itália        | 419               | ≤ 17       | ≥ 24,6                 |

Fonte: Dados da pesquisa (2026) (contagem total por país; citações por documento calculadas).

A presença dos EUA, do Reino Unido e da Itália entre os dez países mais citados sugere que parte do impacto se faz em articulação com países fora da África, e a baixa média de citações de vários países africanos pode refletir o peso dos 284 documentos de 2025, ainda pouco citados.

#### 4.5 Estrutura conceitual e dinâmica temática

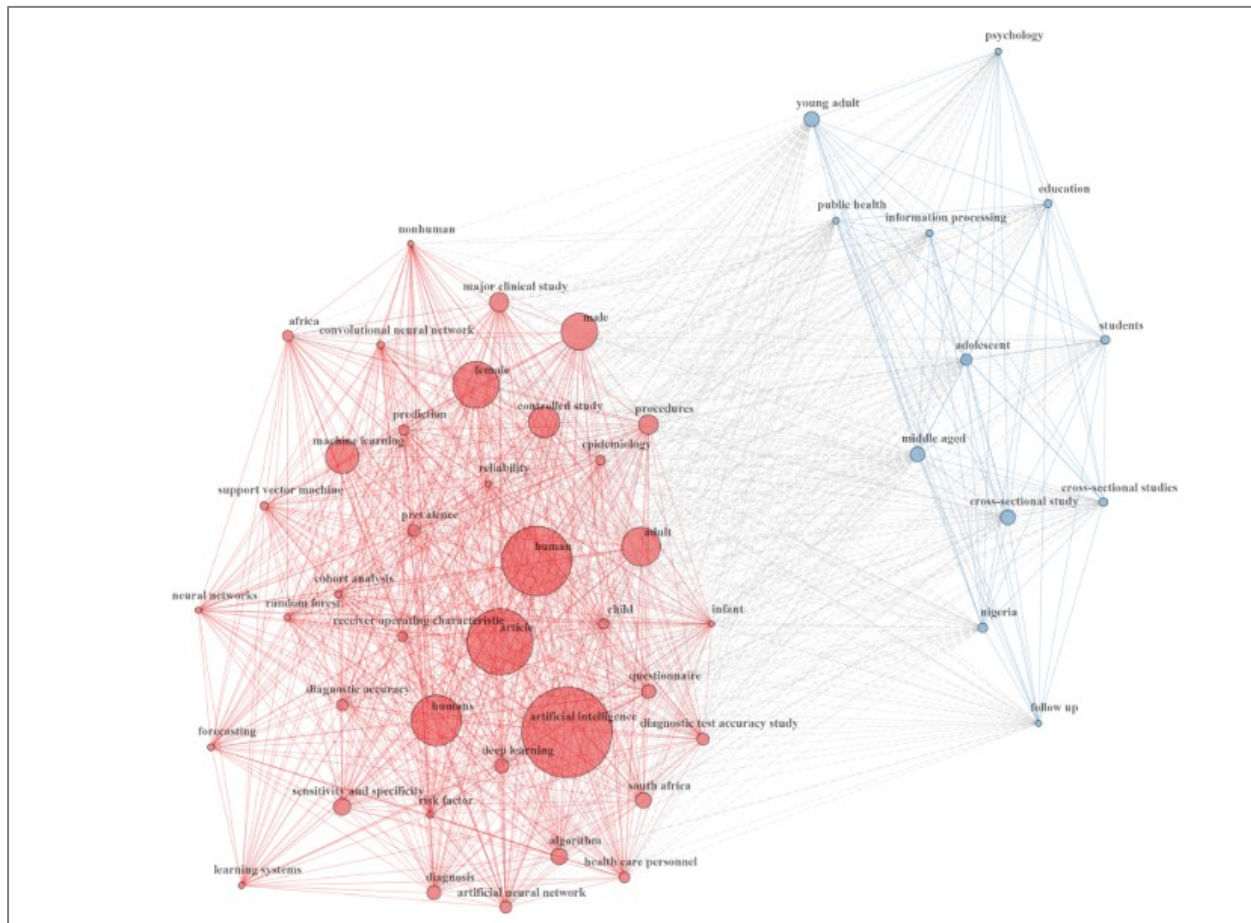
##### 4.5.1 Frequência de palavras-chave

A palavra-chave de autor dominante é artificial intelligence, com 371 ocorrências (59,7% dos documentos), seguida de machine learning (107) e deep learning (23), o que perfaz 501 menções ao núcleo técnico. O segundo bloco é geográfico: Africa (44), South Africa (30), Nigeria (25) e Sub-Saharan Africa (17) somam 116 ocorrências, em coerência com a distribuição por países (Figura 6). Higher education (20) e remote sensing (15) refletem dois dos domínios de maior impacto, a educação e a agricultura e o ambiente. Os sinônimos não foram agregados: a variante AI (13 ocorrências) conta em separado, pelo que a presença efetiva do termo central poderá chegar a 61,8%. A palavra-chave central excede ainda a soma das nove seguintes (294), o que mostra a forte concentração do vocabulário.

#### 4.5.2 Rede de coocorrência e mapa temático

A rede de coocorrência de palavras-chave indexadas (Figura 7) organiza-se em dois clusters (72% e 28% dos nós, respectivamente). O cluster vermelho, dominante, tem como nós centrais artificial intelligence, human, article e humans, e reúne técnicas (machine learning, deep learning, redes neurais, random forest, support vector machine), desenho e desfechos clínicos (controlled study, major clinical study, diagnostic accuracy, sensitivity and specificity) e geografia (africa, south africa). O cluster azul, mais pequeno, agrupa estudos populacionais e de educação, com young adult, adolescent, middle aged, cross-sectional study, education, students, public health e nigeria. As ligações entre os dois clusters são ténues, o que indica uma bifurcação temática entre uma IA técnico-clínica e uma IA ligada a populações e educação.

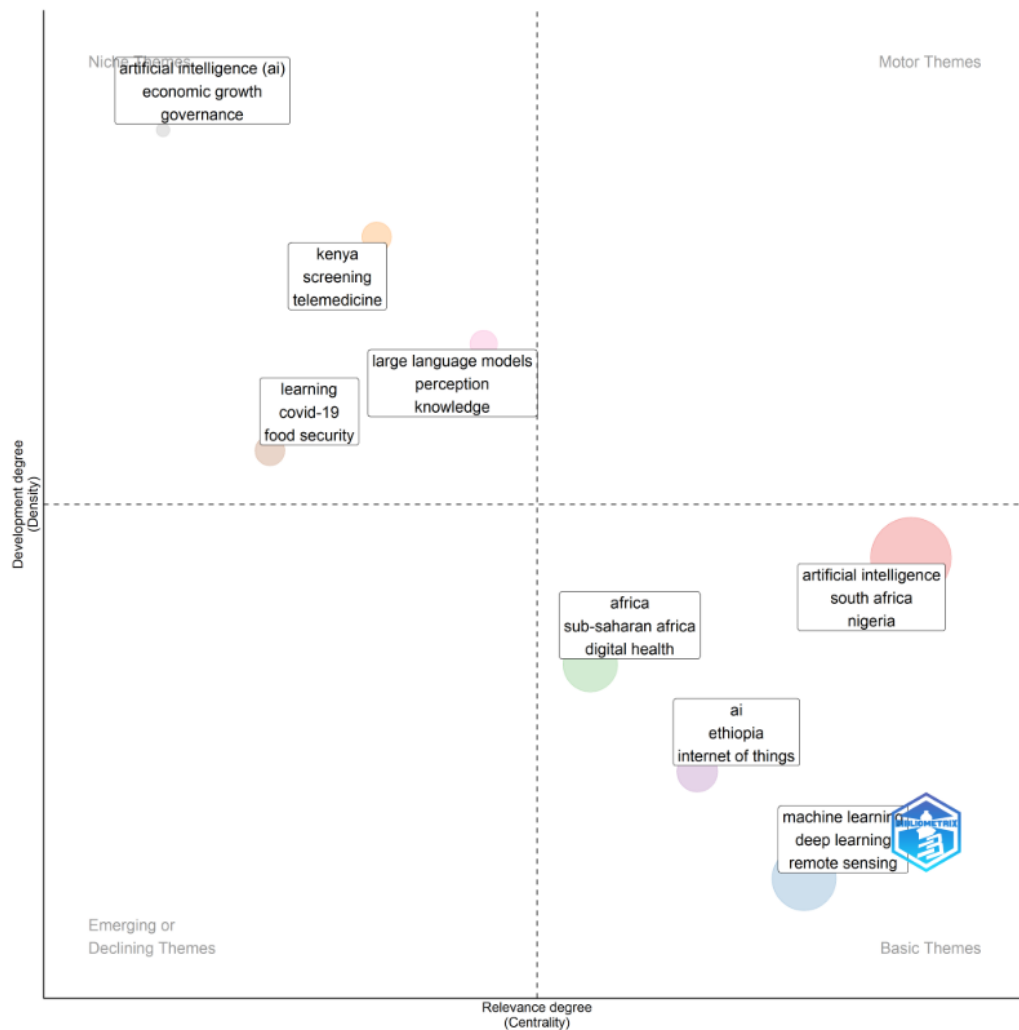
Figura 7 - Rede de coocorrência de palavras-chave indexadas



Fonte: Dados da pesquisa (2026).

O mapa temático (Figura 8) confirma esta estrutura. Os temas básicos, de elevada centralidade e baixa densidade, concentram o *corpus*: *artificial intelligence* / *south africa* / *nigeria* (o maior), *africa* / *sub-saharan africa* / *digital health*, *ai* / *ethiopia* / *internet of things* e *machine learning* / *deep learning* / *remote sensing*. Os temas de nicho, bem desenvolvidos, mas periféricos, são *artificial intelligence (ai)* / *economic growth* / *governance*, *kenya* / *screening* / *telemedicine*, *large language models* / *perception* / *knowledge* e *learning* / *covid-19* / *food security*. Não existem temas motores nem temas emergentes ou em declínio, o que, neste mapa e com estes parâmetros de agrupamento, indica a ausência de temas classificados como motores.

Figura 8 - Mapa temático



Fonte: Dados da pesquisa (2026).

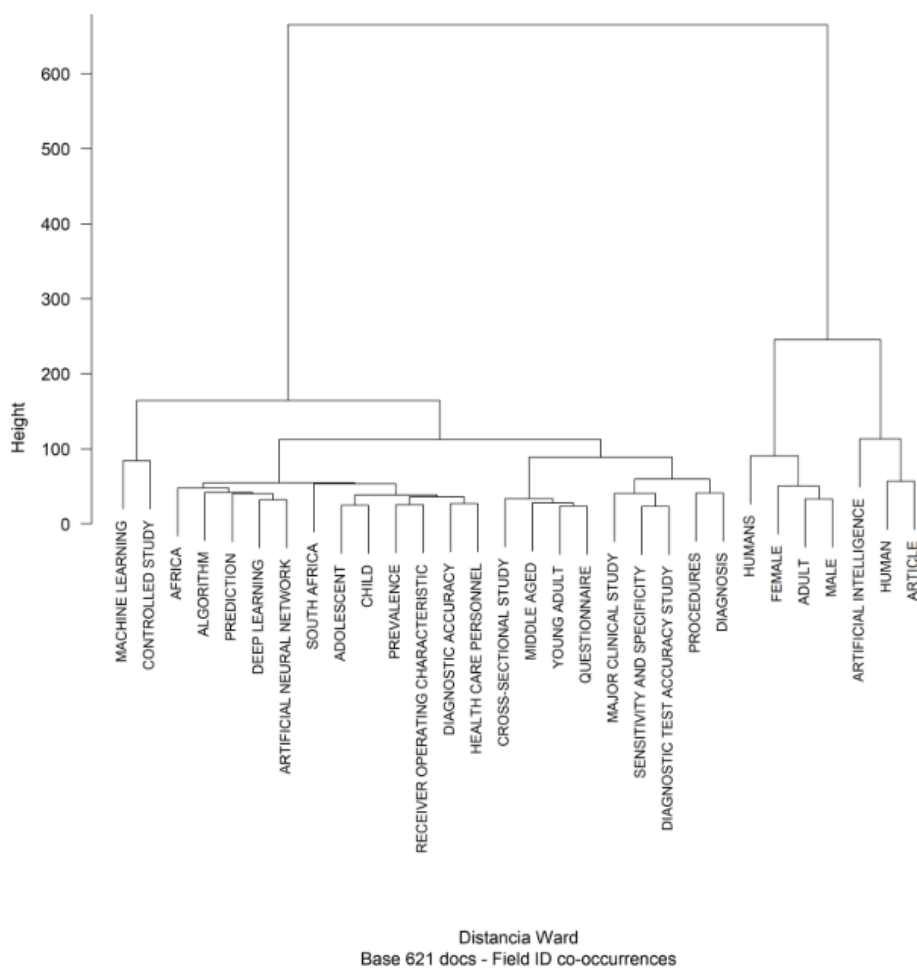
#### 4.5.3 Estrutura fatorial: dendrograma

O dendrograma das 30 palavras-chave indexadas mais frequentes (Figura 9) separa dois macro-ramos a uma altura de cerca de 670. O ramo direito, com sete termos, reúne o vocabulário biomédico genérico (humans, female, adult, male, human, article) e artificial intelligence; funde-se a uma altura de cerca de 245 e as suas fusões internas ocorrem abaixo de cerca de 115. O ramo esquerdo reúne 23 termos de pesquisa aplicada e organiza-se em quatro grupos: (i) um par isolado, machine learning e controlled study (altura  $\approx 85$ ), que só se liga ao restante ramo a cerca de 165;

FILIPE, Arieta Domingas Bartolomeu; DEIXA, Geraldo Vernijo; CHICOTE, Rosalino Subtil; PAULA, Verônica Angélica Freitas de. Evolução e Estrutura da Produção Científica em Inteligência Artificial com Participação de Instituições Africanas (2018–2025): análise bibliométrica de publicações em acesso aberto indexadas na Web of Science e Scopus. *Brazilian Journal of Information Science: research trends*, vol.20, publicação continua, 2026, e026039. DOI: <https://doi.org/10.36311/1981-1640.2026.v20.e026039>.

(ii) um grupo computacional e geográfico, em que Africa, South Africa, algorithm, prediction, deep learning e artificial neural network se associam a termos de saúde (adolescent, child, prevalence, receiver operating characteristic, diagnostic accuracy, health care personnel); (iii) um grupo de estudos transversais (cross-sectional study, middle aged, young adult, questionnaire); e (iv) um grupo de validação diagnóstica (major clinical study, sensitivity and specificity, diagnostic test accuracy study, procedures, diagnosis). Os grupos (iii) e (iv) fundem-se a cerca de 90 e unem-se ao grupo (ii) a cerca de 113.

Figura 9 - Dendrograma da estrutura conceitual (30 palavras-chave indexadas mais frequentes; n = 621)



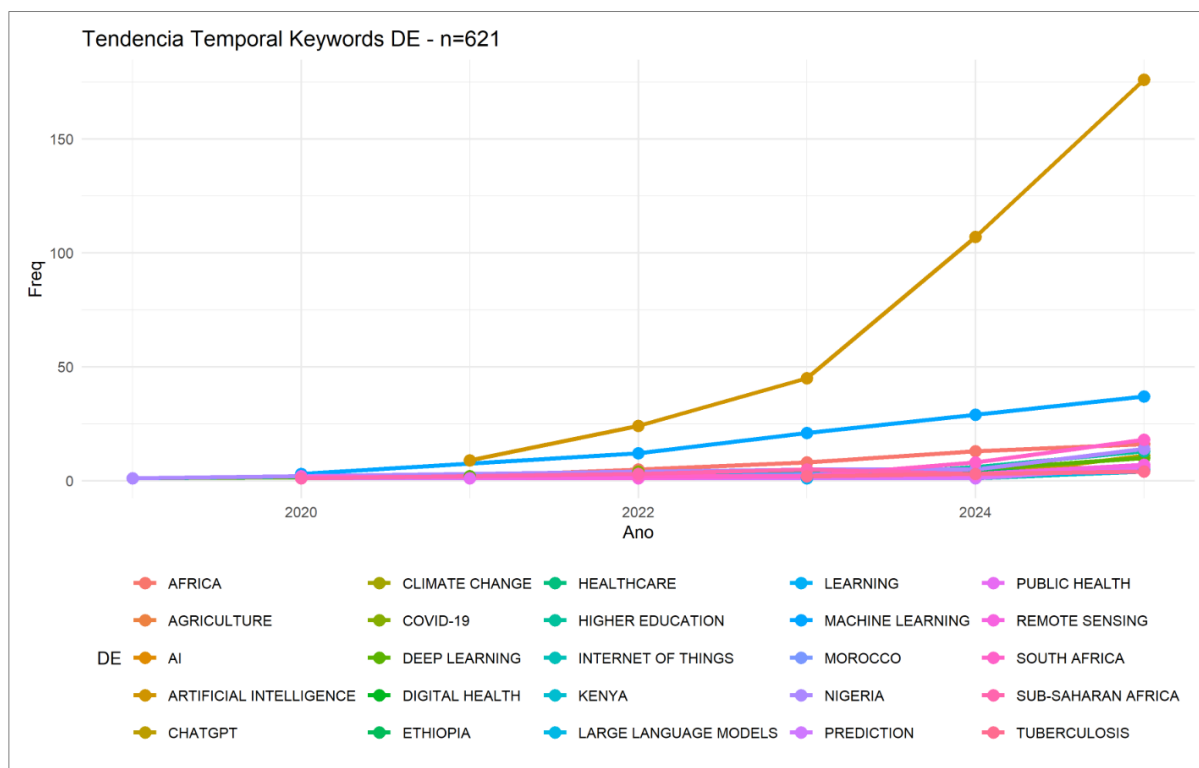
Fonte: Dados da pesquisa (2026).

Os dois macro-ramos são muito dissimilares, o que traduz a coexistência de um vocabulário genérico de indexação, de elevada coocorrência, com subdomínios aplicados (modelagem preditiva, estudos transversais e validação diagnóstica) ainda pouco integrados entre si.

#### 4.5.4 Dinâmica temporal das palavras-chave

A Figura 10 mostra a evolução das palavras-chave de autor. A palavra-chave *artificial intelligence* passa de cerca de 9 ocorrências em 2021 para cerca de 24 em 2022, 45 em 2023, 107 em 2024 e perto de 180 em 2025, uma trajetória exponencial que acompanha a da produção total (Figura 1). A palavra-chave *machine learning* cresce de forma aproximadamente linear, de cerca de 3 ocorrências em 2020 para cerca de 30 em 2024 e 37 em 2025. Todas as restantes palavras-chave permanecem abaixo de cerca de 20 ocorrências em 2025, e termos recentes como *ChatGPT* e *large language models* só podem surgir a partir de 2023 e mantêm frequências baixas.

Figura 10 - Tendência temporal das palavras-chave de autor (DE)



Fonte: Dados da pesquisa (2026).

Em conjunto, a frequência de palavras-chave, a rede, o mapa temático, o dendrograma e a dinâmica temporal descrevem um campo em que a terminologia geral de IA e de aprendizagem automática cresce muito depressa, enquanto os domínios aplicados permanecem dispersos.

## 5 Discussão dos resultados

---

Esta seção interpreta os resultados à luz da literatura bibliométrica e da pesquisa sobre a produção científica na África e discute as suas implicações.

### 5.1 Um campo emergente em fase de crescimento acelerado

---

A trajetória da produção (crescimento anual de 102,99%; 45,7% do corpus em 2025) é compatível com uma fase de expansão do corpus analisado. Price (1963) descreveu o crescimento da ciência como exponencial nas fases iniciais e logístico à medida que se aproxima da saturação, e o ajuste logístico obtido ( $R^2 = 0,999$ ) é coerente com este enquadramento, situando o campo antes do ponto de inflexão estimado (2027,5). É, contudo, necessário relativizar a precisão do ajuste: com oito observações anuais e três parâmetros, a qualidade do ajuste é menos informativa do que a amplitude do intervalo de confiança de K, que não tem limite superior finito. A conclusão limitada ao período observado é que a série anual continua em crescimento e o incremento anual continua a aumentar (de +76 para +124 documentos entre 2024 e 2025), e não a previsão de 1.443 publicações em 2030, que deve ser lida como cenário.

A aceleração a partir de 2022 coincide com a disseminação pública de chatbots baseados em modelos de linguagem (o ChatGPT foi lançado no final de 2022), e os dados sustentam esta associação de forma indireta: a educação é um dos domínios de maior impacto (Abdaljaleel *et al.*, 2024; Essel *et al.*, 2022) e as palavras-chave ChatGPT e large language models surgem apenas nos últimos anos. Trata-se de uma coincidência temporal, e não de uma demonstração causal; o crescimento de artificial intelligence e de machine learning é, aliás, visível já em 2021–2022.

A estrutura temática reforça o diagnóstico de imaturidade. A classificação dos agrupamentos bibliométricos, isoladamente, não permite concluir pela existência ou ausência de

um paradigma científico no sentido de Kuhn (1962). Em termos operacionais, no diagrama estratégico proposto por Callon *et al.* (1991) e usado no mapa temático (Cobo *et al.*, 2011), todos os temas principais são básicos (centrais, mas pouco desenvolvidos) ou de nicho (desenvolvidos, mas periféricos), e não há temas motores.

## 5.2 Dispersão das fontes e transitoriedade autoral

---

A Lei de Bradford (1934) descreve de forma aproximada a dispersão das fontes ( $R^2 = 0,903$ ), mas os quocientes entre zonas não são constantes e o teste de Kolmogorov-Smirnov indica desvio ao modelo, pelo que a lei é validada apenas em sentido qualitativo: um núcleo de 55 fontes (13,1%) produz cerca de um terço do corpus. A ausência de uma revista dominante (a mais produtiva tem 1,6% dos documentos) e a heterogeneidade do perfil das fontes (saúde, educação, detecção remota, direito, gestão) apontam para um campo interdisciplinar, em que a IA funciona como tecnologia transversal aplicada a áreas distintas, e não como um domínio com revistas próprias consolidadas.

Na Lei de Lotka (1926), o expoente clássico é  $\beta \approx 2$ . O valor observado ( $\beta = 4,87$ ; 92,03% de autores com um só documento) e a rejeição do ajuste clássico pelo teste de Kolmogorov-Smirnov ( $D = 0,312$ ) indicam uma comunidade formada sobretudo por autores ocasionais. Esta leitura é consistente com a estrutura temática (domínios aplicados dispersos) e com a média de 6,24 autores por documento: equipes grandes e variáveis, em que a contribuição de cada autor é pontual. Impõem-se duas cautelas. Primeiro, o expoente estimado por regressão log-log é sensível à cauda da distribuição, onde existem apenas um ou dois autores por classe. Segundo, um corpus muito recente inflaciona, por construção, a proporção de autores com um só documento, porque os autores tiveram pouco tempo para voltar a publicar; a transitoriedade pode refletir, em parte, a juventude do campo, e não apenas a sua estrutura social.

## 5.3 Concentração geográfica, colaboração e visibilidade

---

A concentração da produção em poucos países é um traço conhecido da ciência africana. Tijssen (2007) documentou a baixa participação da África na literatura científica mundial e

---

FILIPE, Arieta Domingas Bartolomeu; DEIXA, Geraldo Vernijo; CHICOTE, Rosalino Subtil; PAULA, Verônica Angélica Freitas de. Evolução e Estrutura da Produção Científica em Inteligência Artificial com Participação de Instituições Africanas (2018–2025): análise bibliométrica de publicações em acesso aberto indexadas na Web of Science e Scopus. *Brazilian Journal of Information Science: research trends*, vol.20, publicação continua, 2026, e026039. DOI: <https://doi.org/10.36311/1981-1640.2026.v20.e026039>.

Uthman e Uthman (2007) mostraram que a África do Sul, o Egito e a Nigéria produziram 60% dos artigos biomédicos africanos indexados na PubMed entre 1996 e 2005. A assimetria persiste na IA, onde a África do Sul reúne 43,3% do corpus. A estrutura de colaboração também não é homogênea: Adams *et al.* (2014) mostraram que as redes de colaboração na África formam clusters regionais e ligações externas explicados pela geografia, pela história, pela cultura e pela língua, e não por um sistema global uniforme. A forte presença dos EUA e do Reino Unido (25,4% do corpus em contagem total) e a presença de afiliações de vários países são compatíveis com colaboração internacional, mas não permitem determinar isoladamente a liderança das parcerias ou o seu efeito na visibilidade, o que é compatível com a dissociação entre volume e impacto (Egito, Gana e Itália com médias elevadas; Nigéria com média baixa).

É importante distinguir “produção sobre IA com afiliação africana” de “pesquisa que tem a África como objeto” e de “pesquisa liderada a partir da África”. Entre os dez documentos mais citados, há trabalhos de âmbito global ou multinacional (Defourny *et al.*, 2019; Nti *et al.*, 2020; Abdaljaleel *et al.*, 2024), pelo que a análise bibliométrica, por si só, não permite concluir quem lidera o conhecimento produzido. Katz e Martin (1997) discutem as dificuldades de definir e medir a colaboração, e a medição por afiliações de país (aqui, em contagem total) deve ser complementada com indicadores de liderança, como o país do autor correspondente.

#### 5.4 Domínios de impacto e fragmentação temática

---

Os documentos mais citados mostram que o impacto da IA na África se concentra em aplicações, frequentemente em contextos de recursos limitados: monitoramento agrícola e detecção de doenças de culturas (Defourny *et al.*, 2019; Gomez Selvaraj *et al.*, 2020), rastreio clínico (Belleme *et al.*, 2019), gestão de recursos hídricos (Bhaga *et al.*, 2020; Kouadri *et al.*, 2021) e educação (Abdaljaleel *et al.*, 2024; Essel *et al.*, 2022). Estes domínios correspondem aos dois clusters da rede de coocorrência, um técnico-clínico e outro orientado para populações e educação, e à bifurcação revelada pelo dendrograma, onde um vocabulário genérico de indexação biomédica se separa de subdomínios aplicados. A fragmentação sugere que a pesquisa progride por aplicações paralelas, com pouca integração teórica ou metodológica.

Os temas de nicho do mapa temático (economic growth / governance e large language models / perception / knowledge) indicam que as dimensões de governança e de adoção organizacional estão ainda pouco articuladas com o núcleo técnico. Esta lacuna é relevante porque as barreiras à adoção não são apenas técnicas: Booyse e Scheepers (2024) identificam, na África do Sul, obstáculos sociais, regulatórios e de confiança e transparência, e Achieng e Ogundaini (2024) reportam barreiras à aplicação da análise de dados na vigilância de saúde pública na África Subsaariana. Um campo que avança sobretudo em desempenho algorítmico pode, assim, subestimar os fatores de implementação.

## 6 Conclusões

---

O corpus de publicações sobre IA com afiliação africana representa uma produção recente, com crescimento observado até 2025, sem evidência de saturação no período, produzido por equipes grandes e variáveis, concentrado em poucos países e organizado em subdomínios aplicados pouco integrados. O principal desafio científico não parece ser a escassez de aplicações, mas a consolidação: autores recorrentes, redes de colaboração mais equilibradas e temas motores que articulem desempenho técnico, governança e implementação.

Quanto à contribuição e ao valor, o estudo atualiza o mapeamento bibliométrico da IA na África até 2025, período que inclui a disseminação dos modelos de linguagem de grande porte e que os estudos anteriores não cobrem (Kondo; Diwani, 2023). No plano teórico, integra em um só quadro o modelo de ciclo de vida, o teste das leis de Lotka e de Bradford e a estrutura conceitual, e mostra que um campo emergente e aplicado se afasta da formulação clássica de Lotka e só se ajusta de forma aproximada à de Bradford. No plano prático, o diagnóstico da concentração geográfica, da participação de afiliações internacionais e da fragmentação temática oferece a decisores políticos e a financiadores indicações para apoiar equipes recorrentes, parcerias intra-africanas e pesquisa sobre governança e implementação.

O estudo tem três limitações principais. (1) O corpus reúne apenas documentos revisados por pares, em acesso aberto, indexados na Web of Science e na Scopus, pelo que outras tipologias,

outras fontes e a literatura publicada fora destas bases não estão representadas. (2) O impacto depende da janela de citação, que penaliza os documentos de 2024–2025 (71,5% do corpus), e o ajuste logístico assenta em apenas oito observações anuais. (3) A análise é descritiva e quantitativa: não inclui análise de conteúdo dos documentos, o nível institucional, métricas topológicas das redes de colaboração nem análise de cocitação, e os sinónimos de palavras-chave não foram agregados. Além disso, a filtragem por idioma permanece sem descrição completa, e a estratégia de busca restringe a recuperação de trabalhos que não utilizem os dois descritores seleccionados.

Para minimizar estas limitações, adotaram-se as seguintes estratégias. Quanto à primeira, combinaram-se duas bases e descritores amplos, removeram-se duplicados e documentou-se a seleção segundo o PRISMA 2020. Quanto à segunda, as projeções do modelo logístico foram tratadas como cenários exploratórios, com intervalos de confiança, e contextualizadas por indicadores descritivos (idade média, concentração em 2025 e classificação temática), que não validam independentemente o modelo de crescimento, e o impacto foi reportado por ano de publicação. Quanto à terceira, a estrutura conceitual foi triangulada por cinco técnicas complementares (frequência, rede de coocorrência, mapa temático, dendrograma e tendência temporal), os indicadores foram verificados por consistência interna e os resultados institucionais e de coautoria internacional, que não resistiram a essa verificação, foram excluídos.

Recomenda-se, em trabalhos futuros: (1) complementar a análise com a estrutura intelectual (cocitação e acoplamento bibliográfico) e com a evolução temática por subperíodos, por exemplo 2018–2021, 2022–2023 e 2024–2025; (2) analisar as redes de colaboração nos níveis de autores, instituições e países, com indicadores de liderança, como o país do autor correspondente; e (3) realizar uma análise de conteúdo dos documentos mais citados, centrada na governança de dados, na ética e na validação local das soluções de IA.

## Referências

---

- ABDALJALEEL, M., *et al.* A Multinational Study on the Factors Influencing University Students' Attitudes and Usage of ChatGPT. *Scientific Reports*, v. 14, n. 1, p. 1983, 2024. DOI: <https://doi.org/10.1038/s41598-024-52549-8>.
- ACHIENG, M. S.; OGUNDAINI, O. O. Big Data Analytics for Integrated Infectious Disease Surveillance in Sub-Saharan Africa. *South African Journal of Information Management*, v. 26, n. 1, art. a1668, 2024. DOI: <https://doi.org/10.4102/sajim.v26i1.1668>.
- ADAMS, J., *et al.* International Collaboration Clusters in Africa. *Scientometrics*, v. 98, p. 547-556, 2014. DOI: <https://doi.org/10.1007/s11192-013-1060-2>.
- ALARAN, M. A., *et al.* Challenges and Opportunities of Artificial Intelligence in African Health Space. *Digital Health*, v. 11, 2025. DOI: <https://doi.org/10.1177/20552076241305915>.
- AL FARARNI, K., *et al.* Hybrid Recommender System for Tourism Based on Big Data and AI: A Conceptual Framework. *Big Data Mining and Analytics*, v. 4, n. 1, p. 47-55, 2021. DOI: <https://doi.org/10.26599/BDMA.2020.9020015>.
- ARIA, M.; CUCCURULLO, C. Bibliometrix: An R-Tool for Comprehensive Science Mapping Analysis. *Journal of Informetrics*, v. 11, n. 4, p. 959-975, 2017. DOI: <https://doi.org/10.1016/j.joi.2017.08.007>.
- BELLEMO, V., *et al.* Artificial Intelligence Using Deep Learning to Screen for Referable and Vision-Threatening Diabetic Retinopathy in Africa: A Clinical Validation Study. *The Lancet Digital Health*, v. 1, n. 1, 2019, p. e35-e44, DOI: [https://doi.org/10.1016/S2589-7500\(19\)30004-4](https://doi.org/10.1016/S2589-7500(19)30004-4).
- BHAGA, T. D., *et al.* Impacts of Climate Variability and Drought on Surface Water Resources in Sub-Saharan Africa Using Remote Sensing: A Review. *Remote Sensing*, v. 12, n. 24, art. 4184, 2020. DOI: <https://doi.org/10.3390/rs12244184>.
- BOOYSE, D.; SCHEEPERS, C. B. Barriers to Adopting Automated Organisational Decision-Making through the Use of Artificial Intelligence. *Management Research Review*, v. 47, n. 1, p. 64-85, 2024. DOI: <https://doi.org/10.1108/MRR-09-2021-0701>.
- BRADFORD, S. C. Sources of Information on Specific Subjects 1934. *Engineering*, v. 137, p. 85-86.
- CALLON, M., *et al.* Co-Word Analysis as a Tool for Describing the Network of Interactions between Basic and Technological Research: The Case of Polymer Chemistry. *Scientometrics*, v. 22, p. 155-205, 1991. DOI: <https://doi.org/10.1007/BF02019280>.
- COBO, M. J., *et al.* An Approach for Detecting, Quantifying, and Visualizing the Evolution of a Research Field: A Practical Application to the Fuzzy Sets Theory Field. *Journal of Informetrics*, v. 5, n. 1, p. 146-166, 2011. DOI: <https://doi.org/10.1016/j.joi.2010.10.002>.

- DEFOURNY, P., *et al.* Near Real-Time Agriculture Monitoring at National Scale at Parcel Resolution: Performance Assessment of the Sen2-Agri Automated System in Various Cropping Systems around the World. *Remote Sensing of Environment*, v. 221, p. 551-568, 2019. DOI: <https://doi.org/10.1016/j.rse.2018.11.007>.
- DONTHU, N., *et al.* How to Conduct a Bibliometric Analysis: An Overview and Guidelines. *Journal of Business Research*, v. 133, p. 285-296, 2021. DOI: <https://doi.org/10.1016/j.jbusres.2021.04.070>.
- ESSEL, H. B., *et al.* The Impact of a Virtual Teaching Assistant (Chatbot) on Students' Learning in Ghanaian Higher Education. *International Journal of Educational Technology in Higher Education*, v. 19, n. 1, art. 57, 2022. DOI: <https://doi.org/10.1186/s41239-022-00362-6>.
- GARFIELD, E. Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas. *Science*, v. 122, n. 3159, p. 108-111, 1955. DOI: <https://doi.org/10.1126/science.122.3159.108>.
- GARFIELD, E. Citation Analysis as a Tool in Journal Evaluation. *Science*, v. 178, n. 4060, p. 471-479, 1972. DOI: <https://doi.org/10.1126/science.178.4060.471>.
- GIKUNDA, K. Harnessing Artificial Intelligence for Sustainable Agricultural Development in Africa: Opportunities, Challenges, and Impact. arXiv, 2024, <https://arxiv.org/abs/2401.06171>. Acesso em: 8 out. 2026.
- GOMEZ SELVARAJ, M., *et al.* Detection of Banana Plants and Their Major Diseases through Aerial Images and Machine Learning Methods: A Case Study in DR Congo and Republic of Benin. *ISPRS Journal of Photogrammetry and Remote Sensing*, v. 169, p. 110-124, 2020. DOI: <https://doi.org/10.1016/j.isprsjprs.2020.08.025>.
- KAPLAN, A.; HAENLEIN, M. Siri, Siri, in My Hand: Who's the Fairest in the Land? On the Interpretations, Illustrations, and Implications of Artificial Intelligence. *Business Horizons*, v. 62, n. 1, p. 15-25, 2019. DOI: <https://doi.org/10.1016/j.bushor.2018.08.004>.
- KATZ, J. S.; MARTIN, B. R. What Is Research Collaboration?. *Research Policy*, v. 26, n. 1, p. 1-18, 1997. DOI: [https://doi.org/10.1016/S0048-7333\(96\)00917-1](https://doi.org/10.1016/S0048-7333(96)00917-1).
- KONDO, T. S.; DIWANI, S. A. Artificial Intelligence in Africa: A Bibliometric Analysis from 2013 to 2022. *Discover Artificial Intelligence*, v. 3, n. 1, p. 1-14, 2023. DOI: <https://doi.org/10.1007/s44163-023-00084-2>.
- KOUADRI, S., *et al.* Performance of Machine Learning Methods in Predicting Water Quality Index Based on Irregular Data Set: Application on Illizi Region (Algerian Southeast). *Applied Water Science*, v. 11, n. 12, art. 190, 2021. DOI: <https://doi.org/10.1007/s13201-021-01528-9>.
- KUHN, T. S. *The Structure of Scientific Revolutions*. University of Chicago Press, 1962.

- LOTKA, A. J. The Frequency Distribution of Scientific Productivity. *Journal of the Washington Academy of Sciences*, v. 16, n. 12, p. 317-323, 1926.
- LUCKIN, R., *et al.* *Intelligence Unleashed: An Argument for AI in Education*. Pearson Education, 2016.
- MCCARTHY, J. What Is Artificial Intelligence? Stanford University, 2007, <https://www-formal.stanford.edu/jmc/whatisai/>. Acesso em: 8 out. 2026.
- MIENYE, I. D., *et al.* Artificial Intelligence and Sustainable Development in Africa: A Comprehensive Review. *Machine Learning with Applications*, v. 18, art. 100591, 2024. DOI: <https://doi.org/10.1016/j.mlwa.2024.100591>.
- NILSSON, N. J. *The Quest for Artificial Intelligence: A History of Ideas and Achievements*. Cambridge University Press, 2010.
- NTI, I. K., *et al.* A Comprehensive Evaluation of Ensemble Learning for Stock-Market Prediction. *Journal of Big Data*, v. 7, n. 1, art. 20, 2020. DOI: <https://doi.org/10.1186/s40537-020-00299-5>.
- PAGE, M. J., *et al.* The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. *BMJ*, v. 372, art. n71, 2021. DOI: <https://doi.org/10.1136/bmj.n71>.
- PRICE, D. J. de S. *Little Science, Big Science*. Columbia University Press, 1963.
- PRITCHARD, A. Statistical Bibliography or Bibliometrics?. *Journal of Documentation*, v. 25, n. 4, p. 348-349, 1969.
- RUSSELL, S.; NORVIG, P. *Artificial Intelligence: A Modern Approach*. 4. ed., Pearson, 2021.
- SMALL, H. Co-Citation in the Scientific Literature: A New Measure of the Relationship between Two Documents. *Journal of the American Society for Information Science*, v. 24, n. 4, p. 265-269, 1973. DOI: <https://doi.org/10.1002/asi.4630240406>.
- TIJSSEN, R. J. W. Africa's Contribution to the Worldwide Research Literature: New Analytical Perspectives, Trends, and Performance Indicators. *Scientometrics*, v. 71, n. 2, p. 303-327, 2007. DOI: <https://doi.org/10.1007/s11192-007-1658-3>.
- TRANFIELD, D., *et al.* Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review. *British Journal of Management*, v. 14, n. 3, p. 207-222, 2003. DOI: <https://doi.org/10.1111/1467-8551.00375>.
- UTHMAN, O. A.; UTHMAN, M. B. Geography of Africa Biomedical Publications: An Analysis of 1996–2005 PubMed Papers. *International Journal of Health Geographics*, v. 6, art. 46, 2007. DOI: <https://doi.org/10.1186/1476-072X-6-46>.
- ZIPF, G. K. *Human Behavior and the Principle of Least Effort*. Addison-Wesley, 1949.
- ZUPIC, I.; ČATER, T. Bibliometric Methods in Management and Organization. *Organizational Research Methods*, v. 18, n. 3, p. 429-472, 2015. DOI: <https://doi.org/10.1177/1094428114562629>.

## Dados da pesquisa

---

Os registros bibliográficos exportados da Web of Science e da Scopus, a string de busca (Apêndice A) e os procedimentos de análise estarão disponíveis mediante solicitação aos autores correspondentes.

## Apêndice A - String de busca

---

("artificial intelligence" OR "machine learning") AND (Africa OR Algeria OR Angola OR Benin OR Botswana OR "Burkina Faso" OR Burundi OR "Cabo Verde" OR Cameroon OR "Central African Republic" OR Chad OR Comoros OR Congo OR "Democratic Republic of the Congo" OR "DR Congo" OR "Republic of the Congo" OR "Côte d'Ivoire" OR "Ivory Coast" OR Djibouti OR Egypt OR "Equatorial Guinea" OR Eritrea OR Eswatini OR Ethiopia OR Gabon OR Gambia OR Ghana OR Guinea OR "Guinea-Bissau" OR Kenya OR Lesotho OR Liberia OR Libya OR Madagascar OR Malawi OR Mali OR Mauritania OR Mauritius OR Morocco OR Mozambique OR Namibia OR Niger OR Nigeria OR Rwanda OR "Sao Tome and Principe" OR Senegal OR Seychelles OR "Sierra Leone" OR Somalia OR "South Africa" OR "South Sudan" OR Sudan OR Tanzania OR Togo OR Tunisia OR Uganda OR Zambia OR Zimbabwe)

---

Copyright: © 2026 FILIPE, Arieta Domingas Bartolomeu; DEIXA, Geraldo Vernijo; CHICOTE, Rosalino Subtil; PAULA, Verônica Angélica Freitas de. This is an open-access article distributed under the terms of the Creative Commons CC Attribution-ShareAlike (CC BY-SA), which permits use, distribution, and reproduction in any medium, under the identical terms, and provided the original author and source are credited.

---

Submetido: 29/07/2026

Aceito: 08/10/2026