Preprints.org

Article

Not peer-reviewed version

A Scientometric Analysis of Artificial Intelligence Literature for Engineering Problems

Mario Jiménez Benítez * , Fabio López Pires , Eustaquio Martínez Jara

Posted Date: 25 November 2025

doi: 10.20944/preprints202511.1933.v1

Keywords: artificial intelligence; engineering; scientometrics; bibliometrics; literature review

Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 November 2025 doi:10.20944/preprints202511.1933.v1

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

A Scientometric Analysis of Artificial Intelligence Literature for Engineering Problems

Mario Jiménez Benítez [1,*], Fabio López Pires [2] and Eustaquio Martínez Jara [1]

1 Universidad Nacional del Este
2 Universidad Nacional de Asunción
* Correspondence: mariojb@fpune.edu.py

Abstract

The expansion of Artificial Intelligence (AI) research has generated a massive and complex scientific ecosystem that requires systematic characterization, where no comprehensive studies have analyzed applications for engineering. This work conducts one of the most extensive scientometric analyses to date, encompassing 159,139 publications of the specialized literature indexed in the Web of Science (2005–2024). Using data cleaning, citation normalization (NCII), institutional productivity measures and keyword mining algorithms, the study maps the global evolution of AI research. Results reveal the dominance of Engineering and Computer Science disciplines, with China and the United States leading scientific output. High-impact open-access journals, such as IEEE Access, serve as the main dissemination channels. Emerging topics such as ChatGPT, Big Data, Internet of Things (IoT), and Digital Twins define the current research frontiers. The study provides a macroscopic evidence-based framework for understanding the dynamics of AI research for engineering problems and identifies future directions such as sentiment-based analytics, predictive modeling, and the evaluation of Large Language Models (LLMs) in scientific production. Overall, the main findings highlight AI's growing role as a multidisciplinary driver of innovation across global research ecosystems.

Keywords: artificial intelligence; engineering; scientometrics; bibliometrics; literature review

---

1. Introduction

The exponential growth of Artificial Intelligence (AI) research over the past decades has transformed it into one of the most prolific and interdisciplinary scientific domains in the world. AI has become a foundational enabler of innovation across science, from computer vision and natural language processing to data-driven modelling and automation. However, this expansion has generated an increasingly complex and fragmented research landscape, making it difficult to assess the overall structure, evolution, and impact of the field.

Scientometrics, derived from a combination of "science" and "metrics," is a quantitative study of the production, dissemination, and impact of scientific knowledge. It applies statistical and computational methods to analyze large-scale bibliographic data, uncovering patterns of collaboration, research trends, and the evolution of disciplines. Together with bibliometrics, this provides a rigorous framework for examining scientific ecosystems and evaluating their dynamics over time. Scientometric and bibliometric studies have emerged as valuable tools for systematically examining the dynamics of large scientific ecosystems.

These approaches enable the identification of trends, collaborations, and research frontiers by analyzing publication data, citation networks, and keyword evolution at scale. Several studies have explored AI-related research using these methods, providing insights into productivity patterns, country-level contributions, or the rise of particular subfields such as deep learning and natural language processing.

In recent years, multiple scientometric studies have analyzed the development of AI within specific domains. For example, Darko *et al.* [1] examined AI research in the Architecture, Engineering,

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 November 2025
doi:10.20944/preprints202511.1933.v1

and Construction (AEC) industry; Mustak *et al.* [2] focused on AI applications in marketing; and Jiménez-Gómez *et al.* [3] mapped its role in library and information sciences. Similarly, Yaqoob *et al.* [4] and Njei *et al.* [5] explored the healthcare domain, while other studies investigated AI integration in sensors and embedded systems [6] or examined specific regional contexts such as Africa [5]. Although these studies provide valuable insights, they are limited to their respective fields of application or to narrow disciplinary scopes. To date, no large-scale scientometric analysis has systematically examined AI research applied to engineering problems as a unified domain.

Understanding how AI interacts with engineering research is critical for several reasons. It reveals how computational intelligence is being translated into practical, domain-oriented innovation, provides evidence on which areas are leading or lagging in technological adoption and helps identify interdisciplinary bridges that are shaping new engineering paradigms. However, to the best of our knowledge, a comprehensive scientometric characterization of AI applied to engineering has not been conducted to date.

To address this research opportunity, this study performs one of the most extensive scientometric analyses to date, covering 159,139 AI-related publications indexed in the Web of Science (2005–2024). Considering the large volume of processed information, this study applies data-cleaning techniques, citation normalization (NCII), institutional productivity measures, and advanced keyword-mining algorithms to examine the global evolution of AI research with a special emphasis on engineering applications.

This work is guided by the following **Research Questions (RQs):**

- **RQ1:** *What ware the main patterns of scientific production in AI research applied to engineering disciplines between 2005 and 2024?*
  This question examines global publication statistics, distribution by document type, authorship, and country-level productivity. It provides insights into the multidisciplinary nature of AI research when applied to engineering and identifies how the field has evolved across different academic disciplines and institutional contexts.
- **RQ2:** *How have thematic trends and emerging technologies evolved at this intersection over time?*
  This analysis focuses on the temporal evolution of keywords and research topics using co-occurrence and trend detection techniques. It highlights dominant and emerging themes–such as machine learning, digital twins, and IoT–and describes how engineering-related AI research has diversified over the last two decades.
- **RQ3:** *Which countries, institutions, and authors have led global production, and what collaboration networks define this research domain?*
  This question explored the geographic and institutional distribution of AI research in engineering. It analyzes collaboration structures among countries and organizations to identify regional leadership, co-authorship networks, and centers of excellence that shape the field.
- **RQ4:** *What are the most influential works and topics that shape the frontiers of AI applications in engineering?*
  Finally, this analysis identified the most cited documents, influential authors, and highly visible journals within the domain. It provides an overview of citation dynamics and highlights key contributions that define the knowledge foundations and frontier areas of the AI-driven engineering research field.

By addressing these four RQs, the present study makes the following three main contributions:

1. A large-scale mapping of AI research through the analysis of 159,139 publications using citation normalization and keyword mining methods provides an evidence-based picture of the field's evolution.
2. The identification of thematic clusters and temporal trends linking AI concepts with engineering domains reveal how emerging technologies such as digital twins, IoT, and machine learning are integrated into engineering contexts.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 November 2025 doi:10.20944/preprints202511.1933.v1

3. A domain-specific evaluation of research impact and collaboration networks offer a structured quantitative foundation for future studies on AI-driven engineering research.

The remainder of this paper is organized as follows: Section 2 describes the methodological framework, Section 3 presents an overview of global AI production and explores thematic trends and knowledge networks, Section 4 analyzes scientific impact and productivity and Section 5 focuses on AI research applied to engineering disciplines. Finally, Section 6 summarizes the key findings of this study and discusses future directions.

2. Materials and Methods

A large collection of publications and citations constitutes the basis for scientometric analysis in specific research areas [7], such as Artificial Intelligence. As mentioned above, this study aims to capture the largest number of peer-reviewed articles on AI. Other databases were consulted, with the Web of Sciences database being the largest contributor of publications compared to other databases. The combined number of publications from other journals represents a much smaller number compared to the total Web of Sciences. The goal was to obtain empirical evidence to support the scientific findings of this scientometric study.

In this context, data mining plays an essential role in bibliometric studies, as it enables the efficient analysis of large volumes of scientific and academic information, one of the big data V's (volume). Through advanced processing techniques applied with the *Bibliometrix*[1] package for the R statistical programming language, patterns, correlations and trends in publications can be identified. This approach not only contributes to the detection of emerging research areas but also facilitates the evaluation of the impact of authors, institutions, and journals, adds value to the discovery of hidden relationships in scientific production, and constitutes a fundamental tool for transforming bibliographic data into structured knowledge, optimizing the understanding of scientific dynamics [8].

The main stages of the methodological framework for data collection and cleaning, data processing and correction, productivity and impact metrics keyword analysis and thematic mapping are described below.

2.1. Data Collection and Cleaning

The dataset was retrieved primarily from the *Web of Science* (WoS) Core Collection, which provides highly curated and standardized bibliographic metadata. While other databases were initially examined, WoS offered the broadest and most consistent coverage of the topic under study, representing the most suitable source for large-scale scientometric evaluation.

To capture the largest possible universe of AI-related publications, the general query term `"artificial intelligence*"` was applied to the *topic* field, which included the title, abstract, author keywords, and keyword plus. The wildcard asterisk (`*`) expands the search to all the derived expressions of the term. This query returned 161,384 publications for the period 2005–2024.

Subsequently, data cleaning was conducted to detect and remove incomplete or duplicate records (e.g., missing author names, titles, or affiliations). The resulting dataset comprised 159,139 unique publications, of which 113,541 were peer-reviewed journal articles. The dataset included 4,479,520 cited references, 221,942 author keywords and 346,622 authors, corresponding to an average of 13.84 citations per article. These values form the empirical foundation of the meta-analysis assessment performed in this study.

2.2. Data Processing and Correction

All the retrieved metadata described in Section 2.1 (i.e. authors, titles, affiliations, keywords, document types, and references) were processed and normalized using the *Bibliometrix* package for

[1] https://www.bibliometrix.org

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 November 2025 doi:10.20944/preprints202511.1933.v1

the R statistical environment. This tool enables large-scale bibliographic analysis and visualization through *biblioAnalysis* and *biblioNetwork*.

Given the high volume of data and complexity of scientific bibliographic networks, the procedure also incorporates advanced data-mining and network-analysis techniques. For instance, term extraction from titles/abstracts enables the detection of emerging topics, while co-word, co-authorship and co-citation matrices support multiple-correspondence analysis and network reduction methods. These approaches reveal the conceptual, social and intellectual structures of the field. Although the WoS indexing format is highly standardized, inconsistencies have been detected in author or institution names (e.g., multiple variants of the same organization). To ensure the reliability of the results, an automated validation and correction routine based on string matching, deduplication and normalization heuristics was executed. The cleaned dataset provides a consistent and reproducible basis for subsequent scientometric analyses, and is currently available online[2].

2.3. Productivity and Impact Metrics

Research productivity was analyzed considering well established scientometric indicators, including (1) direct counting, (2) author position, (3) equal credit methods and (4) the equal credit method [7,9]. Among these four productivity metrics, the *equal credit method* assigns fractional credit to each author according to the number of co-authors—was selected as the most appropriate for this large-scale analysis. The normalized page size approach was excluded, as it may bias results for longer manuscripts, regardless of contribution.

Research impact was measured using citation-based indicators. The *Normalized Citation Impact Index* (NCII) presented in Equation 1 was considered to account for publication age and field normalization, thereby allowing comparison across different years and research areas [10].

$$\text{NCII} = \frac{\text{Number of citations per publication}}{\text{Publication longevity (in years)}} \tag{1}$$

2.4. Keyword Analysis and Thematic Mapping

To identify emerging topics and conceptual structures within AI research, keyword co-occurrence networks and thematic mapping were performed. Given the scale of the dataset, advanced text-mining procedures were implemented within the *Bibliometrix* environment to extract and structure conceptual information from titles, abstracts, and keywords. These procedures include tokenization and stemming to normalize terms, computation of keyword co-occurrence matrices, and application of TF–IDF weighting schemes to identify salient concepts. The resulting matrices were analyzed through Multiple Correspondence Analysis (MCA) and community-detection algorithms (Louvain method) to cluster semantically related terms and reveal the conceptual and thematic structure of AI research. This approach allowed for the identification of research fronts and emerging topics over time, enabling the subsequent thematic and trend analyses presented in Section 3.5.

2.5. Summary of Methodological Stages

The methodological workflow presented in the previous sub-sections follows a structured, four-stage process designed to ensure reproducibility and analytical depth.

1. **Data Collection:** Comprehensive retrieval of publications indexed in the *Web of Science* Core Collection (2005-2024) using the topic query `"artificial intelligence*"` applied to titles, abstracts, and keywords.
2. **Data Cleaning:** Systematic detection and elimination of duplicate, incomplete, or inconsistent records. Author names, institutional affiliations, and keywords were standardized through automated matching and normalization routines to enhance metadata integrity in a clean, consistent and reproducible dataset.

[2] https://github.com/mariomjimenez-py/scientometric-AI

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 November 2025 doi:10.20944/preprints202511.1933.v1

3. **Data Processing:** Integration and pre-processing of bibliographic metadata using the *Bibliometrix* package in the R environment. This stage included text-mining procedures such as tokenization, stemming, and term-weighting (TF–IDF) and the generation of co-occurrence matrices for authors, institutions, and keywords. Multiple Correspondence Analysis (MCA) and community detection (Louvain algorithm) were employed to identify latent conceptual structures.
4. **Analysis and Visualization:** Quantitative assessment of research productivity and impact through citation indicators (e.g., NCII) and network-based visualization of collaboration, co-citation, and thematic relationships. Thematic mapping and temporal overlay analysis were used to trace the evolution of research topics and emerging fronts in AI-related engineering research.

This multi-stage pipeline combines scientometric, bibliometric, and data-mining approaches to provide reproducible, evidence-based mapping of the AI field. The integration of statistical normalization, network analysis, and conceptual modeling ensures that the findings presented in the subsequent sections are robust and methodologically transparent.

3. Global Scientometric Landscape Of AI

The following sub-sections introduce the main findings for AI on Disciplinary Distribution, Geographic and Institutional Patterns and Publication Typology and Sources.

3.1. Disciplinary Distribution

A scientometric analysis of academic disciplines was employed to understand the general structure and development of the AI research field. Each publication was categorized according to the classification provided by the Web of Sciences based on the selected publication medium, which may belong to more than one thematic research area.

The distribution of publication data revealed some noteworthy patterns. Although engineering and computer science research areas each have the largest number of contributions related to AI, there are other areas where research on this technology is being conducted.

Moreover, the data presented in Table 1 reveal that most technology-related research remains predominantly concentrated on the development and refinement of the technologies themselves rather than on their broader domain-specific applications. This trend highlights the early-stage maturity of Artificial Intelligence research, which is still oriented toward algorithmic and computational advances.

Table 1. Publications by Research Area.

Research Area	# of Works	% of Works
Engineering	36,535	22.64%
Computer Science	35,970	22.29%
Technology Other Topics	10,977	6.80%
Chemistry	8,864	5.49%
Telecommunications	8,760	5.43%
Materials Science	7,854	4.87%
Physics	7,278	4.51%
Business Economics	6,691	4.15%
General Internal Medicine	6,102	3.78%
Nuclear Medicine Imaging	6,049	3.75%
Environmental Sciences	5,933	3.68%
Neurosciences Neurology	4,163	2.58%
Health Care Services	4,154	2.57%
Oncology	4,134	2.56%
Medical Informatics	3,961	2.45%
Others	3,959	2.45%
Total	**161,384**	**100%**

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 November 2025 doi:10.20944/preprints202511.1933.v1

Looking forward, it will become increasingly essential to investigate the implications and transformative potential of AI across diverse disciplinary perspectives ranging from materials science, physics, and business economics to more specialized domains such as nuclear radiology, neuroscience, and oncology. For instance, in medical imaging analysis, AI-driven diagnostic systems have demonstrated substantial reductions in the time required for clinical evaluation, thereby enhancing the accuracy and efficiency of medical decision-making processes.

3.2. Geographic and Institutional Patterns

To obtain an overview of the contribution patterns, the distribution of publications by country and citation pattern was analyzed. As shown in Table 2, the largest share of AI research was produced by researchers from China (22.50%) and the United States (16.09%), followed by India (5.63%), the United Kingdom (4.51%) and South Korea (4.01%).

Table 2. Publications and Citations by Top 10 Countries.

Country	# of Works	% of Works	Country	# of Citations
China	34,202	21.49%	China	455,931
USA	24,457	15.37%	USA	445,506
India	8,561	5.38%	UK	164,973
UK	6,872	4.32%	India	89,948
S. Korea	6,104	3.84%	Germany	72,354
Germany	5,605	3.52%	S. Korea	69,868
Italy	5,582	3.51%	Australia	69,288
Spain	4,658	2.93%	Italy	62,985
Australia	3,862	2.43%	Canada	61,559
Canada	3,813	2.40%	Iran	51,966

However, these figures alone do not fully reflect the significance of country contributions; a ranking of countries by the most cited publications is also produced. The numbers on the right side of Table 2 show that China leads in total citations with 455,744, followed by the United States with 445,449 citations, the United Kingdom with 164,972, India with 89,903, and Germany with 72,371. Germany also appears among the contributing countries, but in the sixth place.

This pattern may reflect a global race to lead the development and integration of Artificial Intelligence across practical domains, aiming for its optimal application in everyday life. As a result, AI has become embedded in numerous consumer technologies, from smartphones and autonomous vehicles to smart appliances capable of suggesting recipes or automatically placing grocery orders based on their contents.

3.3. Publication Typology and Sources

The choice of publication medium significantly influences both the visibility and scholarly impact of research output. Therefore, it is relevant to examine the types of media preferred by researchers to communicate research findings to the scientific community of AI.

Since the publication (document) type is specified for all AI records retrieved from the Web of Science, it is possible to analyze the overall distribution of publication formats within this field. As shown in Table 3, the majority of AI outputs, on average 81.60%— were published as journal articles. This predominance is noteworthy, as conferences are often assumed to be the primary venue for disseminating cutting-edge technological research. However, journals typically reach a broader and more enduring readership than specialized conferences [9].

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Table 3. Number of Publications by Type.

Type/Period	2005–2010	2011–2015	2016–2020	2021–2024	Total
Publications	3,262	4,174	25,209	128,739	161,384
Journal Article	85.56%	89.53%	75.64%	75.65%	81.60%
Procedure Document	6.59%	2.23%	1.31%	–	2.53%
Article in Review	3.95%	–	11.94%	14.77%	7.67%
Editorial	2.67%	3.70%	6.22%	4.69%	4.32%
Conference Summary	0.80%	0.70%	4.44%	3.04%	2.25%
Book Review	0.37%	0.41%	0.35%	0.25%	0.35%
Others	0.06%	3.43%	0.10%	1.60%	1.30%
Total	100.00%	100.00%	100.00%	100.00%	100.00%

To extend this analysis, the distribution of publications across conferences and journals is examined in detail. Conference papers, although rapidly published, provide a limited temporal and audience scope and are often constrained to event participants. Conversely, journals offer more rigorous peer review processes and greater dissemination potential, despite longer publication times.

Within this dataset, the number of articles per journal is considerably higher than the number of papers published in conferences, underscoring a structural preference for journal-based communication in the AI research field.

Regarding publication venues, *IEEE Access* stands out as the leading outlet, with 3,519 articles, followed by *Applied Sciences (Basel)* with 2,019 articles, as shown in Table 4. These two journals exemplify the increasing tendency of open-access platforms to dominate AI dissemination. Notably, the list of top publication sources reflects the interdisciplinary nature of AI; beyond technology-oriented outlets such as *Electronics* or *Expert Systems with Applications*, high publication activity is also found in applied domains such as energy (*Energies*) and medical sciences (*Cureus Journal of Medical Science*, 642 articles). This distribution demonstrates AI's integration of AI across technological, scientific, and clinical contexts, reinforcing its role as a cross-cutting discipline.

Table 4. Number of Publications per Top 10 Journals.

Source	Articles
IEEE Access	3,517
Applied Sciences (Basel)	2,018
Sensors	1,880
Scientific Reports	1,490
Sustainability	1,345
Electronics	1,070
Diagnostics	1,008
Expert Syst. w/ Applications	849
Energies	798
Cureus J. of Medical Science	640

3.4. Keyword Evolution

Keywords are efficient resources for representing and classifying the content of articles in abstract and condensed forms. From one perspective, keywords provide the basis for analyzing the themes and key aspects that represent a particular research area. Rapid identification of novel topics can be achieved by analyzing the emergence of keywords over a given period. The scientometric study presented in this work extracted 221,942 author keywords and 77,420 source indexed keywords from the considered dataset. Regarding the average distribution of keywords per publication, it was observed that only between three and six keywords are generally used to attract scientific attention.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

This observation is not specific to AI as publishers often specify the minimum and/or maximum number of keywords per publication.

The rankings of the main keywords with high frequencies are present in Table 5. The results indicate that research activities in the field of Artificial Intelligence are primarily concentrated on technological core itself—topics such as *Artificial Intelligence*, *Machine Learning*, *Deep Learning*, *ChatGPT*, *Big Data*, and the *Internet of Things*. However, the analysis also reveals the emergence of context-driven subjects beyond core technology, including social and global phenomena such as *COVID-19*, whose presence in the literature has grown considerably in recent years. This trend reflects the adaptability of AI research to external challenges and its expanding role in addressing multidisciplinary problems.

Table 5. Top 15 Author Keywords per # of Publications.

Author Keywords	**# of Works**
Artificial Intelligence	58,071
Machine Learning	20,717
Deep Learning	14,974
Artificial Intelligence (AI)	3,412
AI	2,785
ChatGPT	2,689
Neural Networks	2,430
Covid-19	2,341
Internet Of Things	2,077
Big Data	2,076
Artificial Neural Network	1,908
Natural Language Processing	1,867
Convolutional Neural Network	1,819
Feature Extraction	1,743
Optimization	1,710

Figure 1 illustrates the annual evolution of the top 15 keywords associated with AI research between 2005 and 2024. The data reveal a sharp and sustained increase in the occurrence of core technological terms such as *Artificial Intelligence*, *Machine Learning*, and *Deep Learning*.

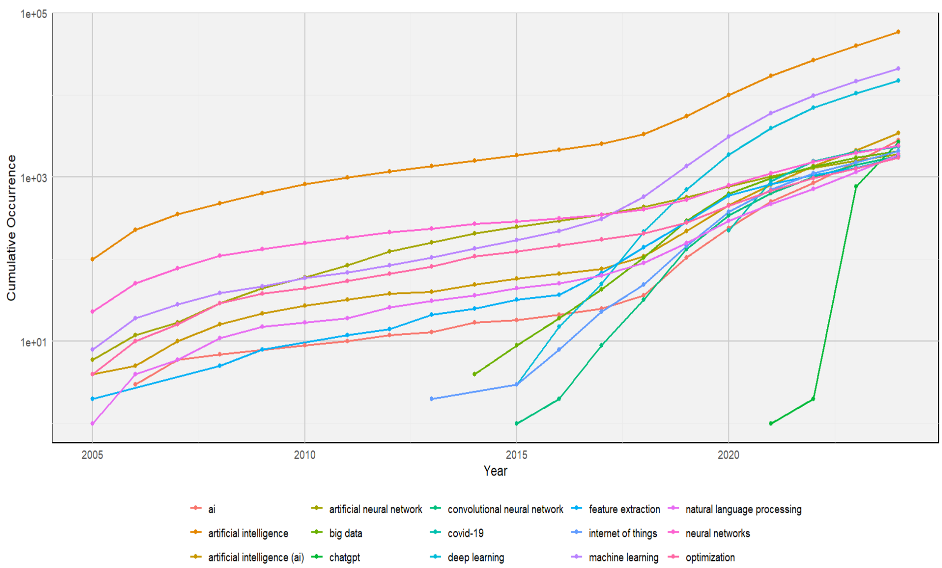

Figure 1. Annual Cumulative Occurrence of Top 15 Author Keywords.

These core technological terms began accelerating markedly from 2019 onward, with the widespread diffusion of AI-driven applications across multiple domains ranging from natural language

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 November 2025 doi:10.20944/preprints202511.1933.v1

processing and computer vision to autonomous systems and predictive analytics marking a transition from theoretical exploration to large-scale technological adoption.

The term *COVID-19* emerged in 2020 as a context-specific keyword that rapidly gained prominence during the global pandemic. Its trajectory reflects how AI methodologies have been repurposed to address urgent societal challenges, particularly in medical diagnostics, epidemiological modeling, and healthcare decision support. This adaptation exemplifies the flexibility of AI in generating rapid data-driven responses under crisis conditions.

Although *COVID-19* appears a with lower overall frequency than foundational concepts, its inclusion underscores the expanding versatility of AI research and its integration into multidisciplinary contexts. Moreover, secondary yet highly correlated terms such as *Prediction*, *Classification*, and *Model* highlight the growing analytical focus on data processing, optimization, and inference across diverse knowledge areas with engineering-specific applications such as the *Internet of Things* or *Big Data*.

Overall, Figure 1 demonstrates the consolidation of Artificial Intelligence and its subfields as the dominant and continually expanding areas of scientific inquiry. The consistent upward trend of key terms over time confirms the centrality of AI in addressing complex, data-intensive problems, while the emergence of new thematic axes indicates its transition toward mature, application-oriented research. Consequently, longitudinal keyword analysis proved to be an effective instrument for detecting emerging topics, understanding thematic evolution, and anticipating future directions for AI research.

3.5. Thematic Clusters and Collaboration Graphs

The thematic map shown in Figure 2—generated in *RStudio* using the *Bibliometrix* environment illustrates the conceptual organization of AI research derived from the keyword co-occurrence networks. The axes represent *Centrality* (horizontal) and *Density* (vertical), which measure the thematic relevance and internal development of the research clusters, respectively. The map divides the field into four quadrants: *Motor*, *Basic*, *Niche*, and *Emerging or Declining Themes*, each corresponding to a phase of maturity.

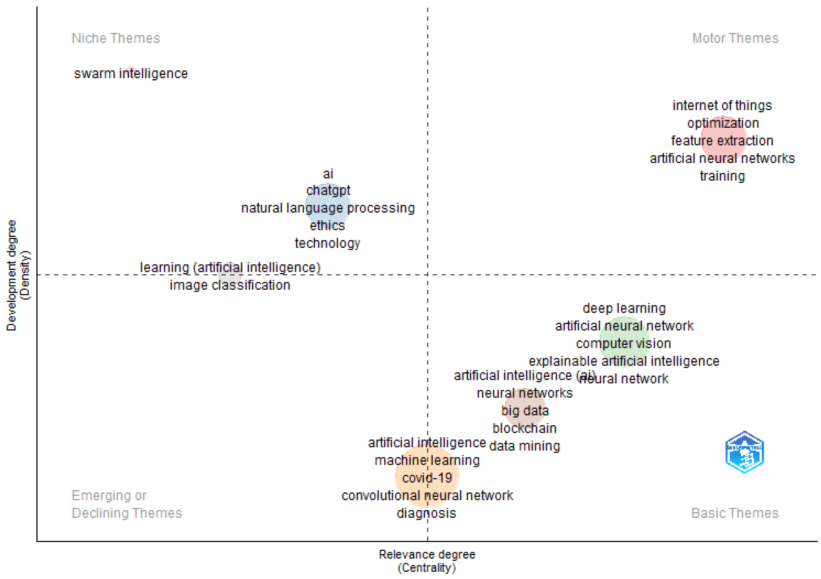

Figure 2. Thematic map of Artificial Intelligence classified by Niche, Emerging, Basic and Motor themes.

In the upper-right quadrant (Motor Themes), highly developed and central topics—such as the *Internet of Things (IoT)*, *Optimization*, *Feature Extraction*, *Artificial Neural Networks*, and *Training*—form the technological core of AI research. These themes exhibited both high relevance and density, indicating strong internal coherence and influence across multiple domains. They represent consolidated areas

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 November 2025 doi:10.20944/preprints202511.1933.v1

of engineering applications and are key drivers for innovation in automation, smart systems, and industrial intelligence.

The lower-right quadrant (Basic Themes) contains fundamental yet broad topics that support the conceptual backbone of the field. Here, terms related to AI such as *Deep Learning*, *Machine Learning*, *Computer Vision*, *Big Data*, *Data Mining*, *Neural Networks*, and *Explainable AI* (XAI) are dominant. These keywords represent the methodological foundation upon which most AI applications are built, serving as transversal enablers for fields including medicine, robotics, and predictive modeling. This centrality suggests that research in these areas will remain influential in shaping the future trajectory of AI.

In the upper-left quadrant (Niche Themes), more specialized and self-contained topics emerge, including *Swarm Intelligence*, which reflects active but narrower lines of investigation based on decentralized, nature-inspired systems. Other keywords such as *Ethics*, *Natural Language Processing*, *Technology*, and *ChatGPT* also appear in this zone, signaling the rise of subfields that are rapidly gaining importance but remain under methodological consolidation. These areas represent potential opportunities for high-impact research as AI expands into socially and linguistically complex domains.

Finally, the lower-left quadrant (Emerging or Declining Themes) groups topics characterized by low centrality and density, such as *Artificial Intelligence (general term)*, *Diagnosis*, and *COVID-19*. While some may correspond to declining interest, others-such as pandemic-related applications—illustrate time-bounded surges of scientific activity triggered by global events. This quadrant also represents the potential areas for future revival or interdisciplinary integration.

Overall, the thematic map demonstrated that AI research simultaneously consolidates mature domains and generates new areas of inquiry. The coexistence of strongly established clusters (e.g., deep learning, IoT) with emerging ones (e.g., ethics, ChatGPT, swarm intelligence) highlights the dynamic and multidisciplinary nature of the field, where technical, ethical, and societal dimensions increasingly converge.

Figure 3 expands the thematic analysis by depicting the temporal evolution of author productivity and its impact within the AI research domain. Each horizontal line represents one of the top 10 leading authors, the bubble size corresponds to the number of published articles (*N. Articles*) and the bubble shade intensity reflects the total citations accumulated per year (*TC per Year*). This visualization highlights both the volume and persistence of research output across nearly two decades (2005–2024), offering a dynamic perspective on the continuity and influence of the field's most prolific contributors.

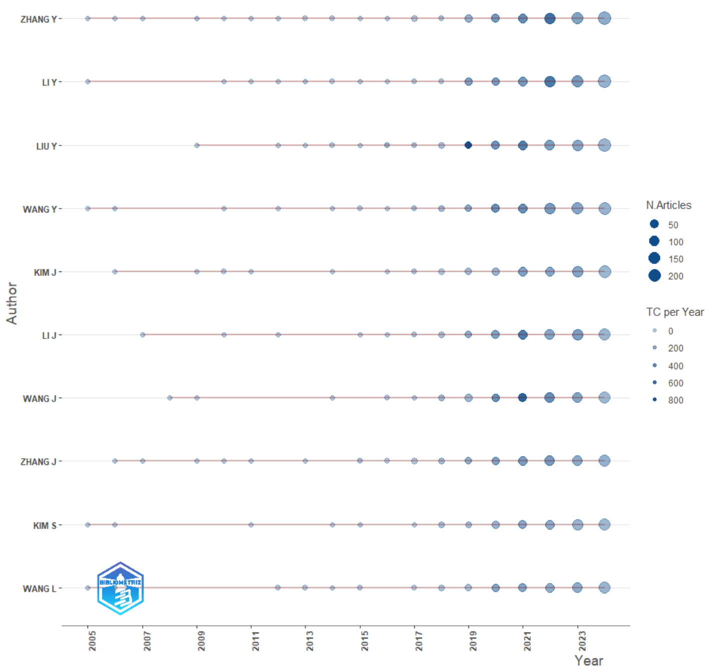

Figure 3. Top 10 Author's Production over Time.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 November 2025 doi:10.20944/preprints202511.1933.v1

The results indicate that authors such as *ZHANG Y.*, *LI Y.*, *LIU Y.*, *WANG Y.*, and *KIM S.* have maintained a sustained and upward trajectory in publication productivity since approximately 2015, with notable acceleration after 2018. Their citation intensity (darker circles) increased significantly during the 2019–2023 period, coinciding with the consolidation of deep learning, neural networks, and optimization techniques as dominant AI research paradigms. This pattern reflects cumulative expertise, in which early contributions provided a foundation for ongoing influence and high visibility in subsequent years. Other researchers, including *LI J.*, *WANG J.*, and *ZHANG J.*, exhibit steady productivity with slightly lower citation density, suggesting a consistent but less concentrated impact compared to the top-tier group. Nonetheless, their longitudinal activity denotes methodological specialization and strong participation in multi-author collaborations, an essential feature of AI's interdisciplinary ecosystem.

The collective output of these authors demonstrates a concentration of intellectual leadership largely within Chinese and East Asian institutions, reinforcing the findings in Table 2, where China emerges as the global leader in both research volume and citation performance. Moreover, the absence of major geographic gaps across the timeline suggests continuous engagement rather than episodic contributions, confirming the sustained expansion of the AI research community over time.

In summary, Figure 3 show a robust pattern of cumulative productivity, where experienced authors consistently contribute to the advancement of Artificial Intelligence through iterative refinement of methods and diversification of applications.

This trajectory underscores the growing institutional maturity and strategic investment in AI research observed over the last decade.

Figure 4 show the temporal evolution of Artificial Intelligence (AI) publications from 2005 to 2024. The curve exhibited a clear exponential growth pattern, with three distinguishable phases of development. During the first phase (2005–2015), the research output remained relatively modest, reflecting the formative years of modern AI dominated by theoretical and algorithmic exploration. A moderate increase began around 2015, coinciding with the resurgence of neural networks and consolidation of deep learning as a dominant paradigm.

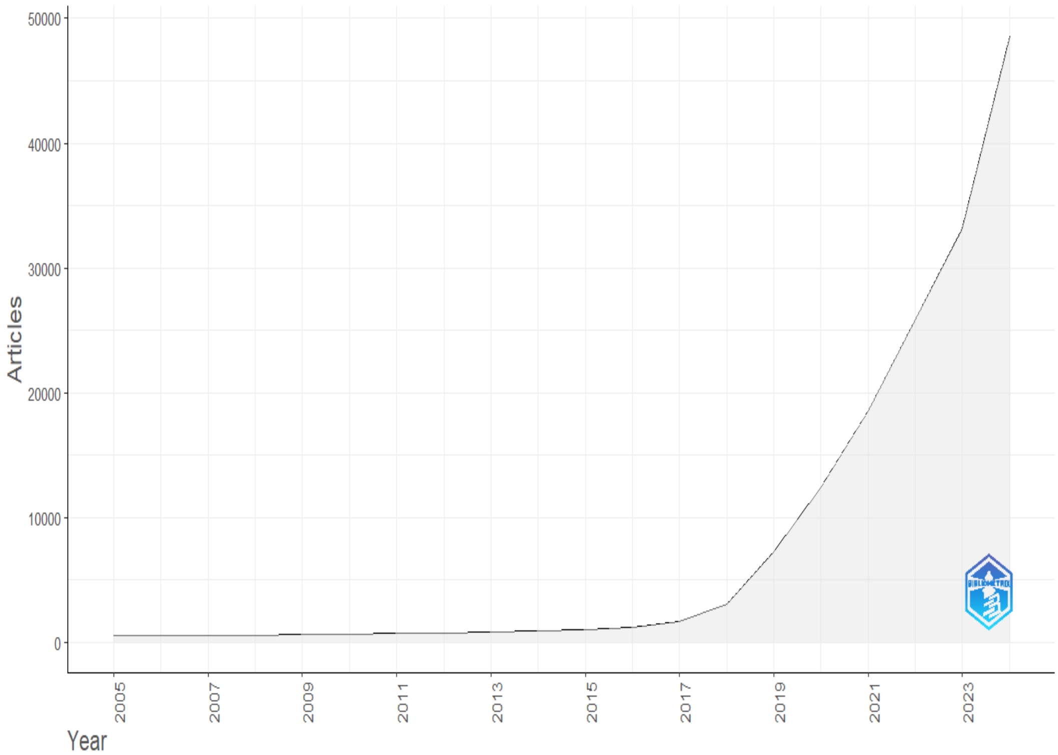

Figure 4. Annual Scientific Output over Time.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 November 2025 doi:10.20944/preprints202511.1933.v1

A pronounced inflection point appeared between 2017 and 2018, marking the transition from steady growth to exponential expansion. This surge aligns with the democratization of high-performance computing resources, the emergence of open-source frameworks such as *TensorFlow* and *PyTorch*, and the rapid adoption of GPUs and cloud-based AI infrastructures. These technological enablers significantly lowered barriers to entry for experimentation and large-scale data processing.

From 2020 onward, the publication rate accelerates sharply, reaching its historical maximum by 2024. This period coincides with the proliferation of generative models, reinforcement learning applications, and the integration of AI into nearly all scientific and engineering domains. The cumulative pattern underscores AI's transition from a specialized research topic to a pervasive scientific discipline driving innovation across sectors.

Overall, the figure reflects a sustained exponential trajectory, suggesting that AI-related publications will continue to expand in terms of volume and thematic diversity. This trend highlights not only the technological maturation of AI but also its growing role as a foundational component of modern scientific inquiry.

4. Scientific Impact and Productivity in AI

Understanding the scientific impact and productivity of AI research complements the structural and thematic evidence presented in Section 3. Here we shift from *what* is being published to *how much* influence those outputs exert and *who* sustains their influence. The section is organized in two parts: (i) *Citation Dynamics*, which quantifies the impact over time using normalized indicators and (ii) *Most Influential Works and Authors*, which identify institutions and researchers that anchor the intellectual hierarchy of global AI.

4.1. Citation Dynamics

To compare the impact across publication years, we relied on the *Normalized Citation Impact Index* (NCII), which standardizes raw citations by years in print [11]. This adjustment mitigates temporal bias and enables cross-period comparison of long-term influence.

Table 6 (Top 15 Publications by Total Citations) reveals three consistent patterns. *(1) Persistence of early baselines:* The average NCII shows a mild downward slope, indicating that foundational contributions have a disproportionate influence over time. *(2) Impact clustering (2016–2020):* The consolidation of deep learning and its rapid diffusion coincide with the most cited window: Emblematic cases include *Generative Adversarial Networks* [12] (31,828 citations; NCII 7,957), *Mastering the Game of Go with Deep Neural Networks and Tree Search* [13] (9,009; NCII 1,126) and *Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks* [14] (6,616; NCII 945). *(3) Thematic breadth of high impact:* Highly cited studies span bio-inspired optimization (*Ant Colony Optimization* [15]; *Artificial Bee Colony* [16,17]), systems/overview pieces (*Machine Learning: Trends, Perspectives, and Prospects* [18]) and governance-oriented themes (*Explainable AI* [19]; *Federated Learning* [20]).

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 November 2025 doi:10.20944/preprints202511.1933.v1

Table 6. Top 15 Publications by Total Citations (TC), Yearly Total Citations (YTC) and Normalized Total Citations (NTC).

Research Work Title and Reference	Author and Journal	DOI	TC	YTC	NTC	NCII
Generative adversarial networks [12]	Goodfellow I, 2020, Communications of the ACM	10.1145/3422622	31,828	5,305	953.0	7,957
Mastering the game of Go with deep neural networks and tree search [13]	Silver D, 2016, Nature	10.1038/nature16961	9,009	901	185.0	1,126
Dermatologist-level classification of skin cancer with deep neural networks [14]	Esteva A, 2017, Nature	10.1038/nature21056	6,616	735	115.6	945
A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm [16]	Karaboga D, 2007, Journal of Global Optimization	10.1007/s10898-007-9149-x	5,045	266	106.9	721
Mastering the game of Go without human knowledge [21]	Silver D, 2017, Nature	10.1038/nature24270	4,887	543	85.4	87
Federated Machine Learning: Concept and Applications [20]	Yang Q, 2019, ACM Transactions on Intelligent Systems and Technology	10.1145/3298981	4,870	696	128.4	974
Machine learning: Trends, perspectives, and prospects [18]	JORDAN MI, 2015, Science	10.1126/science.aaa8415	4,177	380	99.0	835
Computational Radiomics System to Decode the Radiographic Phenotype [22]	Van Griethuysen JJM, 2017, Cancer Research	10.1158/0008-5472.CAN-17-0339	3,457	384	60.4	494
Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI [19]	Arrieta AB, 2020, Information Fusion	10.1016/j.inffus.2019.12.012	3,121	520	93.4	780
Ant colony optimization [15]	Dorigo M, 2006, IEEE Computational Intelligence Magazine	10.1109/MCI.2006.329691	3,077	154	84.1	171
How to conduct a bibliometric analysis: An overview and guidelines [23]	Donthu N, 2021, Journal of Business Research	10.1016/j.jbusres.2021.04.070	2,941	588	143.1	980
Overcoming catastrophic forgetting in neural networks [24]	Kirkpatricka J, 2017, Proceedings of the National Academy of Sciences of the United states of America	10.1073/pnas.1611835114	2,826	314	49.4	404
Highly sensitive flexible pressure sensors with microstructured rubber dielectric layers [25]	Mannsfeld SCB, 2010, Nature Materials	10.1038/nmat2834	2,579	161	62.5	184
On the performance of artificial bee colony (ABC) algorithm [17]	Karaboga D, 2008, Applied Soft Computing	10.1016/j.asoc.2007.05.007	2,526	140	64.4	158
High-performance medicine: the convergence of human and artificial intelligence [26]	Topol EJ, 2019, Nature Medicine	10.1038/s41591-018-0300-7	2,445	349	64.5	489

This dispersion confirms that citation impact in AI is not monolithic but distributed across methodological and application frontiers.

Across the full corpus, citation counts range from single digits to 31,000+, with a mean of 13.84 citations per article and an average annual rate of 2,575, evidencing intense referencing and fast knowledge diffusion. According to document type (see Table 3), journals concentrate ∼81.60% of citations, reflecting their editorial rigor and archival reach, while conferences, books and chapters account for the remainder. Notably, review articles although few are disproportionately cited (e.g., [23]), underscoring their role as synthesis instruments that stabilize emerging agendas in a rapidly evolving field.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 November 2025 doi:10.20944/preprints202511.1933.v1

4.2. Most Influential Works and Authors

While the previous sub-section characterizes *how* impact accumulates, this section details *where* and *by whom* it is generated.

Table 7 ranks organizations by publication volume and share. The *University of California System* leads (2,709; 1.81%), followed by *Harvard University* (1.68%) and the *Chinese Academy of Sciences* (1.67%).

Table 7. Publications by Top 10 Institutions.

Institution	# of Works	% of Works
University of California System	2,709	1.81%
Harvard University	2,510	1.68%
Chinese Academy of Sciences	2,497	1.67%
University of London	2,429	1.62%
Harvard Univ. Medical Affiliates	1,813	1.21%
Egyptian Knowledge Bank (EKB)	1,723	1.15%
University of Texas System	1,413	0.94%
Stanford University	1,405	0.94%
Harvard Medical School	1,401	0.94%
CNRS (France)	1,327	0.89%

The prominence of *Harvard University Medical Affiliates* and *Harvard Medical School* highlights the growing biomedical orientation of AI, while the presence of the *Egyptian Knowledge Bank (EKB)* points to regional open-access infrastructure enabling wider knowledge production. Overall, the distribution mirrors sustained investments in the U.S. and China and the rise of health-related AI.

Table 8 lists the top authors by H-index, G-index and M-index, capturing sustained influence (H), weight of landmark papers (G), and citation velocity adjusted by career length (M). *LI Y.* attains the highest H-index (52) with 9,906 citations (668 papers, since 2005). *WANG J.* leads in G-index (92) and M-index (2.833), indicating faster accumulation of high-impact outputs. *LIU Y.* shows the peak G-index (118) with 15,642 citations, signaling exceptional concentration of influence. The remaining leaders (i.e. *ZHANG Y., WANG Y., ZHANG J., LI J., KIM J., WANG L., KIM S.*) illustrate the strong East-Asian footprint in contemporary AI production.

Table 8. H-index, G-index and M-index of the Top 10 cited authors. PY-Start means the year starting publications.

Author	H-Index	G-Index	M-Index	TC	# of Works	PY-Start
LI Y	52	85	2.47	9,906	668	2,005
WANG J	51	92	2.83	10,346	531	2,008
LIU Y	50	118	2.94	15,642	668	2,009
ZHANG Y	46	87	2.19	10,539	685	2,005
WANG Y	45	85	2.14	9,823	660	2,005
ZHANG J	44	76	2.20	8,067	528	2,006
LI J	43	76	2.26	8,328	537	2,007
KIM J	38	58	1.90	5,443	556	2,006
WANG L	36	62	1.71	5,516	453	2,005
KIM S	32	59	1.52	4,891	487	2,005

Tables 6, 7 and 8 show a mature yet rapidly expanding ecosystem: high-impact works cluster around deep learning and its applications, institutional leadership is concentrated in the U.S. and China with a notable biomedical tilt, and author-level influence is dominated by East-Asian networks.

Considering that in the global scientometric analysis of AI presented in Section 3, the field of *Engineering* has emerged as the most prolific and influential domain, this section presents a more specialized and in-depth examination focused on AI applied to Engineering Problems. While the

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

preceding sections offered a macroscopic overview of AI's evolution across disciplines, the following analysis narrows the scope to characterize the distinctive research patterns, institutional dynamics, and thematic structures that define engineering-oriented AI research. This specialized exploration provides deeper insights into how AI methodologies are developed, disseminated, and applied to engineering problems, revealing both technological and domain-specific trends.

This section is organized as follows: First, an Overview quantifies the engineering-focused subset in terms of publications, references, and authorship, comparing its citation behavior and dissemination patterns to those observed in the broader AI corpus. Next, Impact and Leading Institutions assess geographic distribution, national research performance, and institutional leadership, highlighting the countries and organizations that drive scientific output and citation influence.

Additionally, Collaboration and Domain Trends explore co-authorship networks, keyword co-occurrence structures, and international collaboration patterns, providing insights into how global partnerships and thematic clusters are formed in engineering-oriented AI research. Finally, Thematic Evolution in Engineering traces the temporal progression of core topics, revealing how research interests have consolidated and diversified across different periods.

Together, these subsections offer a refined and longitudinal perspective on how Artificial Intelligence has evolved as a foundational enabler of modern engineering research, illustrating the interplay between methodological innovation, institutional capacity, and global collaboration in shaping one of the most dynamic fields of contemporary science.

4.3. Overview Impact and Leading Institutions

By reprocessing 35,535 documents from the total studied in Section 3, we obtained 1,305,784 references, 79,531 author keywords and 86,486 authors.

The analyzed dataset yields an average of 19.85 citations per article, notably higher than the global average for Artificial Intelligence publications, which stands at 13.65 citations per article. Table 9 shows that most AI studies focused on engineering—approximately 86.10%—are published as journal articles. This trend remains consistent with the general AI publication pattern discussed in Section 3, reinforcing the hypothesis that journals are the preferred medium for disseminating research results to a broad audience, in contrast to conferences, where audience size and reach are typically limited [27,28].

Table 9. Number of Publications by Type in the field of AI for Engineering.

Type/Period	**2005–2010**	**2011–2015**	**2016–2020**	**2021–2024**	**Total**
Publications	965	1,419	5,598	27,553	35,535
Journal Article	84.25%	90.56%	87.41%	82.20%	86.11%
Procedure Document	10.36%	1.90%	1.86%	0.23%	3.59%
Article in Review	2.28%	3.74%	7.23%	11.32%	6.14%
Editorial	–	1.97%	2.75%	1.72%	1.61%
Book Review	2.28%	–	–	0.03%	0.58%
Others	0.83%	1.83%	0.75%	4.49%	1.98%
Total	100.00%	100.00%	100.00%	100.00%	100.00%

5. Artificial Intelligence Applied to Engineering

To better understand the contribution dynamics, the distribution of publications by country and citation count was analyzed. As presented in Table 10, China and the United States lead in research output, accounting for 29.13% and 8.47% of the work, respectively. However, citation indicators provide further context regarding the impact: China ranks first with 198,210 citations, followed by the United States with 76,201 citations. These results confirm the leading role of both countries in shaping AI research in engineering domains.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 November 2025 doi:10.20944/preprints202511.1933.v1

Table 10. Publications and Citations by the Top 10 Countries in the field of AI for Engineering.

Country	# Works	% Works	Country	# Citations
China	10,352	29.13	China	198,210
USA	3,009	8.47	USA	76,201
India	2,480	6.98	India	36,058
Korea	1,916	5.39	UK	34,723
Italy	1,161	3.27	Korea	30,186
UK	1,131	3.18	Iran	25,181
Spain	1,076	3.03	Australia	24,270
Iran	978	2.75	Canada	20,329
Canada	846	2.38	Italy	20,235
Australia	767	2.16	Spain	16,033

Figure 5 illustrates the institutional contribution landscape, showing that the Chinese Academy of Sciences is the most prolific institution, with 1,056 publications, followed by the Egyptian Knowledge Bank (906) and the National Institute of Technology system (559). The remaining institutions were distributed mainly across Asia and North America, mirroring the geographical patterns observed in Table 10.

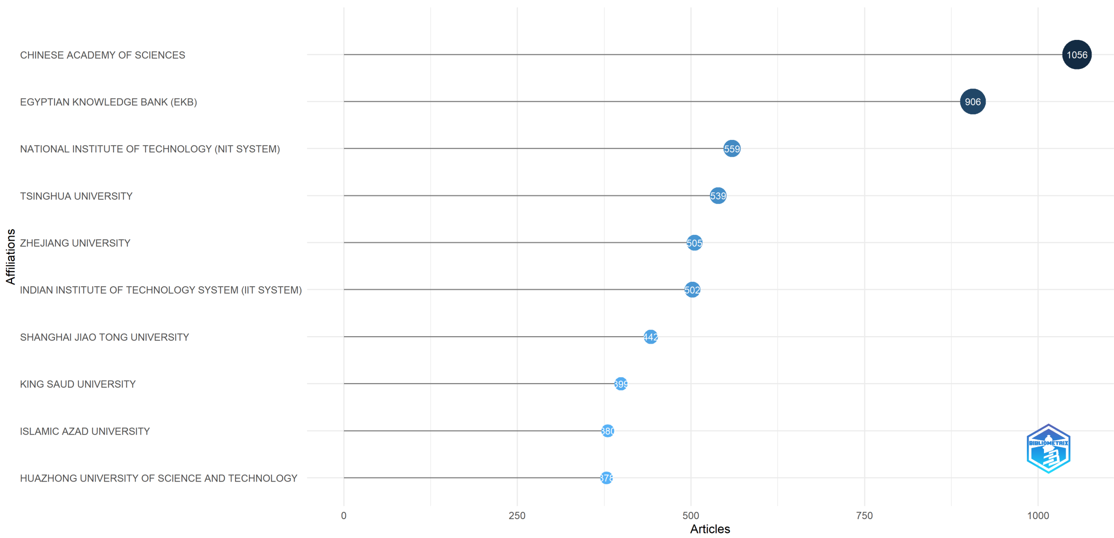

Figure 5. Top Contributing Affiliations by Number of Publications in the field of AI for Engineering.

The analysis of publication venues, summarized in Table 9, confirms the predominance of journal publications over other document types such as conference proceedings, reviews, or book chapters. IEEE Access has emerged as the leading outlet with 3,399 articles, followed by Applied Sciences–Basel with 1,968 publications, as detailed in Table 11.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 November 2025 doi:10.20944/preprints202511.1933.v1

Table 11. Number of Publications per Top 10 Journals in the field of AI for Engineering.

Source	Articles
IEEE Access	3,399
Applied Sciences–Basel	1,968
Sensors	1,846
Electronics	1,053
Expert Systems with Applications	831
Engineering Applications of AI	565
IEEE Internet of Things Journal	501
Multimedia Tools and Applications	431
Wireless Communications	365
Computers in Biology and Medicine	284

Notably, interdisciplinary journals such as Computers in Biology and Medicine also appeared among the top sources, highlighting the increasing integration of AI methods across engineering and biomedical contexts.

This scientometric study identified 79,531 author keywords and 24,551 indexed keywords. The most frequent terms (see Table 12) reflect a strong technical orientation, featuring terms such as machine learning, deep learning, feature extraction, Iot, neural networks, and task analysis.

Table 12. Top 15 Author Keywords by Number of Publications in the field of AI for Engineering.

Author Keywords	# Works
Artificial Intelligence	12,180
Machine Learning	4,938
Deep Learning	4,408
Feature Extraction	1,281
Learning (AI)	1,184
Internet of Things	1,124
Artificial Intelligence (AI)	1,026
Neural Networks	990
Optimization	927
Training	907
Task Analysis	796
Data Models	784
Artificial Neural Network	777
Computer Vision	710
Computational Modeling	662

Among aggregated terms, the most cited are model, prediction, and artificial intelligence. Across both author-defined and indexed keyword sets, optimization and neural networks recur as central concepts, underscoring the emphasis on algorithmic refinement and applied modeling within engineering-oriented AI research.

The average number of citations per document is 19.85, and the average number of citations per document per year is 3.971, with a total of 1,305,784 references used—figures that differ from those in the general AI analysis. Applying the NCII formula to the list of most cited engineering documents, Table 13 shows that the most cited work is [29], *Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)*, with 2,825 citations, an NCII of 471, and an average of 353.13 citations per year.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Table 13. Top Cited Research Articles in the field of AI for Engineering by Total Citations (TC), Yearly Total Citations (YTC) and Normalized Total Citations (NTC).

Paper	Author	DOI	TC	YTC	NTC	NCII
Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI) [29]	Adadi A, 2018	10.1109/ACCESS. 2018.2870052	2,825	353.13	49.9	471
Deep Reinforcement Learning: A Brief Survey[30]	Arulkumaran K, 2017	10.1109/MSP. 2017.2743240	2,314	257.11	43.5	331
Efficient Processing of Deep Neural Networks: A Tutorial and Survey [31]	Sze V, 2017	10.1109/JPROC. 2017.2761740	2,309	256.56	43.4	330
Multimodal Machine Learning: A Survey and Taxonomy [32]	Baltrusaitis T, 2019	10.1109/TPAMI. 2018.2798607	2,002	286.00	40.5	400
The Arithmetic Optimization Algorithm [33]	Abualigah L, 2021	10.1016/j.cma. 2020.113609	1,837	367.40	58.9	612
Applications of machine learning to machine fault diagnosis: A review and roadmap [34]	Lei YG, 2020	10.1016/j.ymssp. 2019.106587	1,678	279.67	44.1	420
Fog and IoT: An Overview of Research Opportunities [35]	Chiang M, 2016	10.1109/JIOT.2016. 2584538	1,539	153.90	29	192
ViBe: A Universal Background Subtraction Algorithm for Video Sequences [36]	Barnich O, 2011	10.1109/TIP. 2010.2101613	1,444	96.27	35.1	111
Artificial intelligence for fault diagnosis of rotating machinery: A review [37]	Liu RN, 2018	10.1016/j.ymssp. 2018.02.016	1,415	176.88	25	236
Automated detection of COVID-19 cases using deep neural networks with X-ray images [38]	Ozturk T, 2020	10.1016/j.compbiomed. 2020.103792	1,373	228.83	36.1	343
Artificial intelligence in healthcare [39]	Yu KH, 2018,	10.1038/s41551-018-0305-z	1,371	171.40	24.2	228
Machine Learning in Agriculture: A Review [40]	Liakos KG, 2018,	10.3390/s18082674	1,293	161.60	22.9	215
Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data [41]	Jia F, 2016,	10.1016/j.ymssp. 2015.10.025	1,291	129.10	24.4	161
LSTM network: a deep learning approach for short-term traffic forecast [42]	Zhao Z, 2017,	10.1049/iet-its. 2016.0208	1,153	128.10	21.7	164
Towards 6G wireless communication networks: vision, enabling technologies, and new paradigm shifts [34]	You XH, 2021,	10.1007/s11432-020-2955-6	1,147	229.40	36.8	382

Another highly cited work is [33], *The Arithmetic Optimization Algorithm*, with an NCII of 612 and 1,837 citations, averaging 367.4 citations per year despite its recent publication. These works stand out as the most influential in engineering-related AI research.

5.1. Collaboration and Domain Trends

Figure 6 show the co-authorship network of AI research applied to engineering. The visualization revealed three major clusters representing distinct but interconnected research communities. Zhang Y led the red cluster, followed by Wang Y, Wang G, Li Y, and Wang H. The green cluster is centered around Kim J, with active collaboration between Kim H, Lee S, and Kim Y. Third, a smaller blue cluster does not have a dominant leader, although Kumar A appears to be more collaborative than Kumar S or Khan M. A. The green cluster exhibits the highest collaboration intensity, whereas the red cluster shows broader cross-group interactions, with multiple links bridging authors from different communities.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 November 2025 doi:10.20944/preprints202511.1933.v1

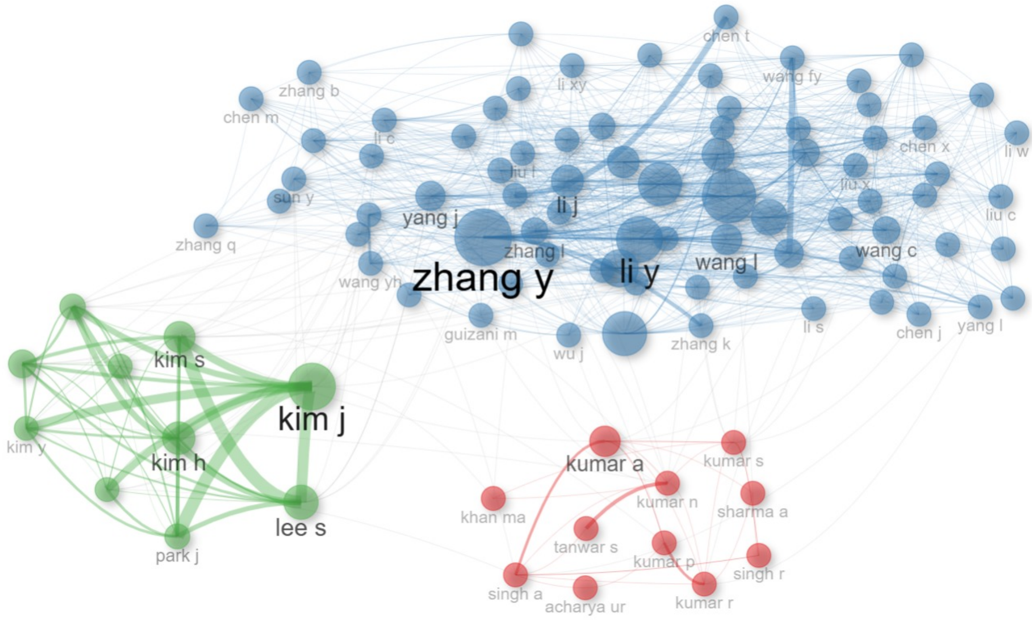

Figure 6. Co-Authorship Collaboration Network in the field of AI for Engineering.

The most central authors—Zhang Y, Wang Y, Li Y, Kim J, and Kim H—emerge as key connectors in the global engineering-oriented AI research network, reflecting both disciplinary concentration and international reach.

Beyond individual collaborations, the network also revealed the role of institutional and regional proximity in shaping research communities.

The predominance of Asian authors, especially from China and South Korea, indicates the strong influence of national research ecosystems, which often promote high-output collaboration networks driven by government-funded initiatives. Conversely, European and North American participation appears more fragmented but still relevant, focusing primarily on interdisciplinary projects linking AI with applied sciences, biomedical engineering, and sustainability. This distribution pattern underscores the growing concentration of engineering-oriented AI expertise in East Asia, while maintaining global knowledge diffusion through cross-continental collaborations.

Figure 7 illustrates the network of author-assigned keywords, based on the 50 most cited terms in the engineering field, revealing three well-defined thematic clusters.

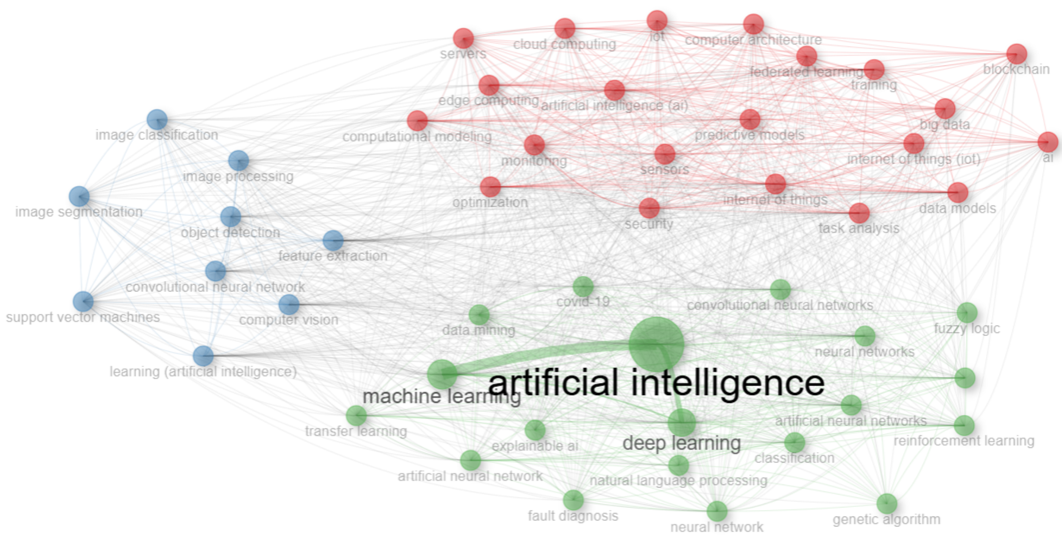

Figure 7. Keyword Co-Occurrence Network in the field of AI for Engineering Research.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 November 2025 doi:10.20944/preprints202511.1933.v1

In the green cluster, *artificial intelligence* appears to be the dominant term, closely associated with *machine learning*, *deep learning*, and *data mining*, which together define the backbone of the field. The blue cluster groups terms related to image-based applications, including *image processing*, *object detection*, and *image segmentation*,thereby highlighting the growing importance of visual computing in engineering contexts. The red cluster aggregates infrastructure- and system-related terms such as *data models*, *sensors*, *security*, and *servers*, suggesting the integration of AI into networked and industrial systems. Inter-cluster linkages indicate a high degree of interdisciplinarity, with recurrent conceptual overlaps among AI, machine learning, and deep learning.

The dominance function implemented in the *Bibliometrix* package [43] was applied to compute the ranking of authors based on their relative contributions as first authors. This metric quantifies leadership by calculating the proportion of articles in which an author appears first relative to their total output, thus reflecting their individual impact within collaborative environments. According to Table 14, Kumar A ranks first with a dominance index of 0.2822, and is the first author in 35 of 125 publications. Wang Y and Li Y appear as first authors in 37 papers, ranking third and fifth respectively, while Zhang Y, despite having the largest total number of publications (193), exhibits a lower dominance factor due to extensive collaboration. This analysis underscores the distributed yet hierarchical nature of authorship leadership across the engineering AI community.

Table 14. Ranking of authors by dominance in AI applied to engineering.

Author	Dominance Factor	Total Articles	Single-Authored	First-Authored	Rank by Articles
Kumar A	0.2822	125	1	35	8
Wang H	0.2377	124	2	29	9
Li Y	0.2356	160	3	37	5
Wang Y	0.2298	165	4	37	3
Wang J	0.2131	128	6	26	7
Kim J	0.2049	163	2	33	4
Liu Y	0.1944	181	1	35	2
Zhang Y	0.1904	193	4	36	1
Zhang J	0.1241	147	2	18	6
Lee S	0.1160	112	0	13	10

From a scientometric perspective, the dominance factor complements traditional metrics such as the H-index or G-index by capturing patterns of intellectual leadership rather than overall productivity. The data reveal a structure in which prolific authors frequently collaborate across institutional boundaries but maintain clear hierarchies within project leadership. This model,which is prevalent in large-scale engineering projects, reflects the organizational nature of AI research, in which interdisciplinary teams combine computational, physical, and systems engineering expertise.

Such configurations tend to accelerate innovation while concentrating on authorship credit within leading groups.

Figure 8 expands the analysis to the country level, revealing intense collaboration between China and the United States as the dominant international axis of engineering AI research. Regional networks extend across Europe, East Asia, and the Middle East, with increasing participation from developing regions. This visual pattern aligns with that in Table 10, confirming China's central role in both production and international cooperation. The geographical clustering of partnerships further suggests that national funding programs, academic exchange networks, and open-access initiatives play a decisive role in shaping the global topology of AI research collaborations.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 November 2025 doi:10.20944/preprints202511.1933.v1

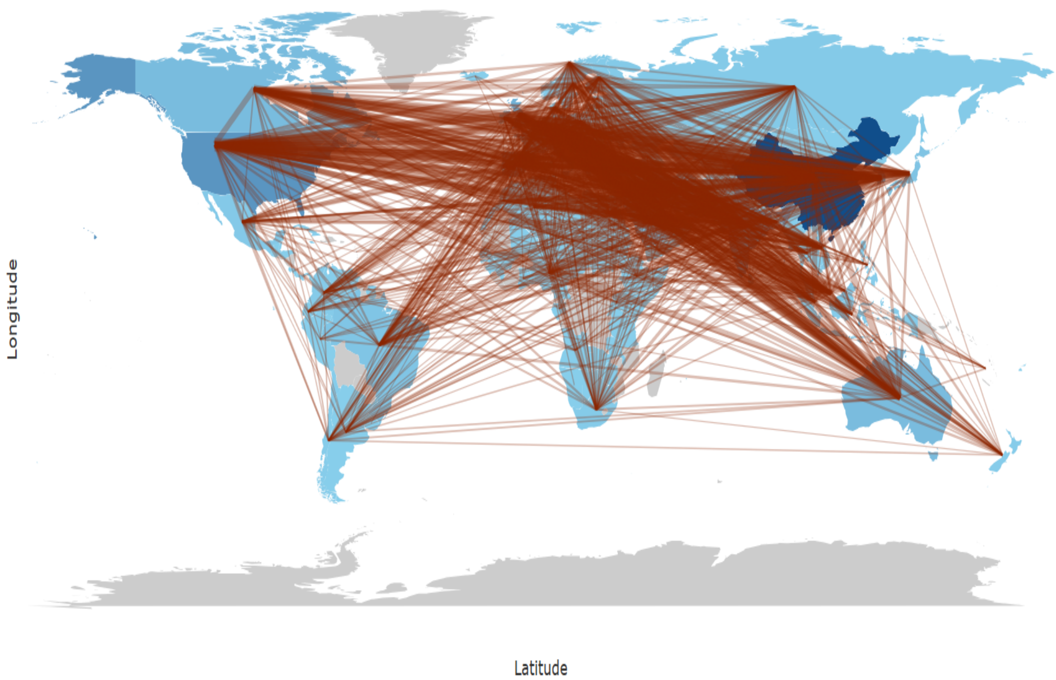

Figure 8. Country Collaboration Network in the field of AI for Engineering Research.

The three-field diagram in Figure 9 provides a multivariate view of the relationships between countries, keywords, and journals. Thicker connecting lines indicate stronger co-occurrence links, revealing how the most productive countries—chiefly China, the United States, and India—are associated with high-frequency keywords such as *machine learning*, *deep learning*, and *optimization*, and with leading publication venues, such as *IEEE Access* and *Applied Sciences–Basel*. This reinforces the interconnectedness of national productivity, thematic specialization, and preferred journals within engineering-oriented AI research. The visualization also highlights emerging countries such as Iran and Egypt, which are gaining visibility through open-access collaboration strategies and regional research consortia, signaling a gradual diversification of the AI research landscape.

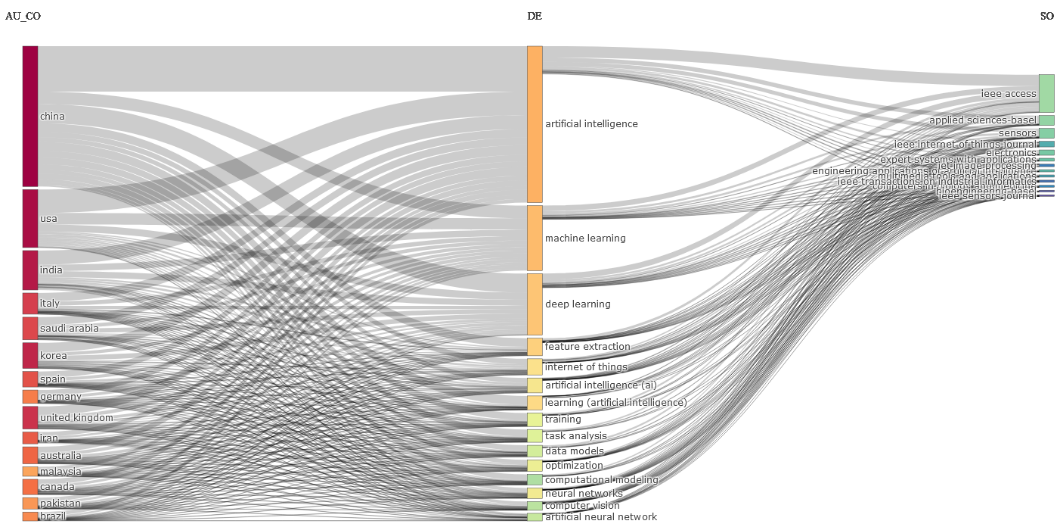

Figure 9. Three-Field Plot of Countries, Author Keywords, and Journals in Engineering.

Finally, these collaborative and thematic trends portray an ecosystem that is increasingly networked, multidisciplinary, and globally distributed. The convergence of machine learning, optimization, and industrial applications suggests that engineering-oriented AI is transitioning from exploratory to applied maturity.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 November 2025 doi:10.20944/preprints202511.1933.v1

Moreover, the leadership patterns and collaboration intensities observed at both author and country levels reveal how knowledge production in engineering AI operates within complex hierarchies of expertise, where localized clusters feed into a broader, interconnected structure of global innovation.

6. Conclusions

This paper presents one of the most comprehensive scientometric analyzes of Artificial Intelligence (AI) research to date, encompassing 159,139 publications indexed in *Web of Science* (2005–2024). By integrating bibliometric, scientometric, and data-mining approaches—including citation normalization (NCII), co-occurrence networks, and thematic mapping this study provides a data-driven understanding of how AI research has evolved, who drives it, and how it is applied across engineering and other scientific disciplines.

The following synthesis outlines the main findings (MFs) derived from this study, organized in relation to the four Research Questions (RQs) defined in Section 1.

6.1. Research Questions and Main Findings

RQ1. What are the main patterns of scientific production in AI research applied in engineering disciplines (2005–2024)?

MF1: AI research output has grown exponentially, from fewer than 4,000 publications in 2005–2010 to more than 128,000 in 2021–2024 (Table 3, Section III.C). This growth indicates a sustained expansion of AI as one of the fastest-growing scientific domains worldwide.

MF2: Engineering (22.64%) and Computer Science (22.29%) were the dominant disciplinary categories in AI-related publications, followed by Technology, Chemistry, and Physics (Table 1, Section III.A), confirming AI's technical core and cross-domain influence of AI.

MF3: The majority of AI research is disseminated through journal articles (81.60%), indicating a structural preference for peer-reviewed journals over conferences (Table 3, Section III.C), even in a rapidly evolving technological field.

MF4: The leading publication outlets are *IEEE Access* and *Applied Sciences (Basel)*, followed by *Sensors* and *Scientific Reports* (Table 4, Section III.C), reinforcing the trend toward open-access, high-visibility platforms for AI dissemination.

MF5: Geographically, China (22.50%) and the United States (16.09%) dominate global AI research output and citations, followed by India, the United Kingdom, and South Korea (Table 2, Section III.B). This concentration defines a bipolar global leadership pattern centered in Asia and North America.

MF6: At the institutional level, the University of California System, Harvard University, and Chinese Academy of Sciences are the most prolific contributors (Table 7, Section IV.B), highlighting the role of large, research-intensive universities in shaping the AI landscape.

RQ2. How have thematic trends and emerging technologies evolved in this intersection over time?

MF7: Keyword evolution (Figure 1, Section III.D) shows sustained growth in *Artificial Intelligence*, *Machine Learning*, and *Deep Learning* since 2015, with accelerated expansion after 2019. This confirms a structural shift from theoretical foundations to applied, data-driven approaches.

MF8: Contextual topics such as *COVID-19* emerged sharply in 2020, reflecting the responsiveness of AI research to global challenges, such as pandemic modeling and diagnostics.

MF9: Thematic mapping (Figure 2, Section III.E) identified seven core clusters dominated by *Optimization*, *Training*, *Internet of Things*, and *Big Data*, indicating the convergence of AI with cyber-physical systems and data engineering.

RQ3. Which countries, institutions, and authors have led global production, and what collaboration networks define this research domain?

MF10: Collaboration networks (Figure 6, Section V.A) reveal three large author clusters, with leading contributors including Zhang Y, Wang Y, Li Y, and Kim J. Collaboration intensity is strongest in Asian institutions, particularly in China and South Korea.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

MF11: Country-level collaboration (Figure 8) shows the strongest bilateral link between China and the United States, followed by Europe–Asia collaborations. Despite global participation, AI research remains geographically concentrated in only a few innovation hubs.

RQ4. What are the most influential studies and topics that shape the frontier of AI applications in engineering?

MF12: The most cited AI papers (Table 6, Section IV.A) include foundational works such as *Generative Adversarial Networks* (Goodfellow et al., 2020) and *Mastering the Game of Go* (Silver et al., 2016). In the engineering subset (Table 13, Section V.A), the most cited are *Explainable Artificial Intelligence (XAI)* (Adadi & Berrada, 2018) and *The Arithmetic Optimization Algorithm* (Abualigah et al., 2021), signaling a thematic shift toward interpretability and algorithmic efficiency.

MF13: The most productive authors—Li Y, Wang J, Liu Y, and Zhang Y, achieved the highest H-index and G-index values (Table 8, Section IV.D), confirming the correlation between long-term productivity and sustained citation impact.

MF14: The Domain Classification (Table 14, Section V.A) reveals that Kumar A and Wang Y exhibit the highest dominance factors, reflecting leadership in first-authorship and domain specialization within engineering-oriented AI research.

MF15: The thematic evolution analysis (Figures A1–A5 in APPENDIX A) demonstrates the chronological transition from algorithmic research (e.g., Genetic Algorithms, Support Vector Regression) to integrated applications such as *IoT*, *Digital Twins*, and *Natural Language Processing*. The latest period (2023–2024) introduces emerging topics such as *Neuromorphic Computing* and *Spiking Neural Networks*.

6.2. Future Works

The empirical evidence presented reveals that AI research has reached the consolidation, maturity, and specialization phases. The transition from generic technological development to domain-oriented innovation (e.g., medicine, materials, and energy) indicates a reconfiguration of engineering paradigms toward data-intensive and predictive systems.

Based on the presented main findings, several future directions are proposed:

- **(F1) Predictive Scientometrics:** Develop time-series and machine learning models to forecast the evolution of AI publication output and thematic shifts.
- **(F2) Sentiment and Semantic Analysis:** Apply natural language processing (NLP) and topic modeling to evaluate tone, polarity, and narrative evolution in AI literature.
- **(F3) Role of LLMs in Scientific Production:** Assess how generative models (e.g., ChatGPT) affect authorship practices, productivity, and citation behaviors.
- **(F4) Open-Access Dynamics:** Quantitatively compare the NCII of open-access journals (e.g., *AI Open Access*) versus subscription-based outlets to evaluate dissemination efficiency.
- **(F5) Interdisciplinary Engineering Applications:** Extend domain-specific scientometric analyzes to subfields such as smart grids, biomedical systems, and autonomous robotics, where AI integration is most intense.

In conclusion, this study offers a reproducible, quantitative, and multi-dimensional mapping of the global AI research landscape. The identified trends, collaborations, and thematic clusters provide a robust foundation for future policy decisions, academic strategies, and technological foresight in AI-driven engineering research

Author Contributions: Conceptualization, M.J.B; methodology and software, M.J.B.; validation and data curation, M.J.B. and F.L.P.; writing original draft preparation, M.J.B and F.L.P.; writing review and editing, M.J.B and F.L.P.; visualization,M.J.B. and E.M.J.; supervision, E.M.J.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 November 2025 doi:10.20944/preprints202511.1933.v1

Informed Consent Statement: Not applicable

Data Availability Statement: The data presented in this article are available on https://github.com/mariomjimenez-py/scientometric-AI (accessed on 19 October 2025).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

NCII	Normalized Citation Impact Index
IoT	Internet of Thing
LLM	Large Language Model
NLP	Natural Language Processing
UK	United Kingdom
MCA	Multiple Correspondence Analysis
TF-IDF	Term frequency-inverse document frequency

Appendix A. Figures on Thematic Evolution

Appendix A.1. Figures, Tables and Schemes

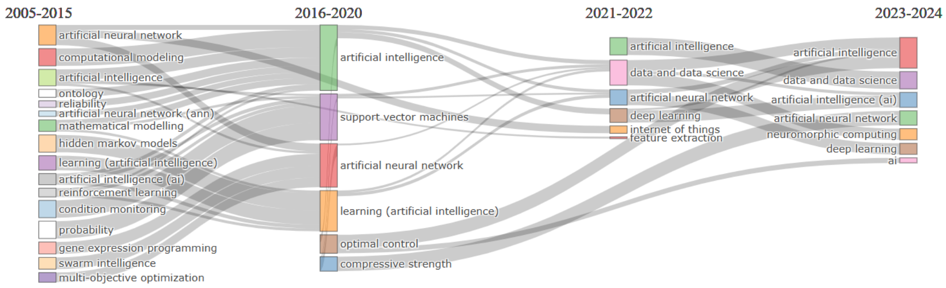

Figure A1. General Thematic Evolution Across Four Periods.

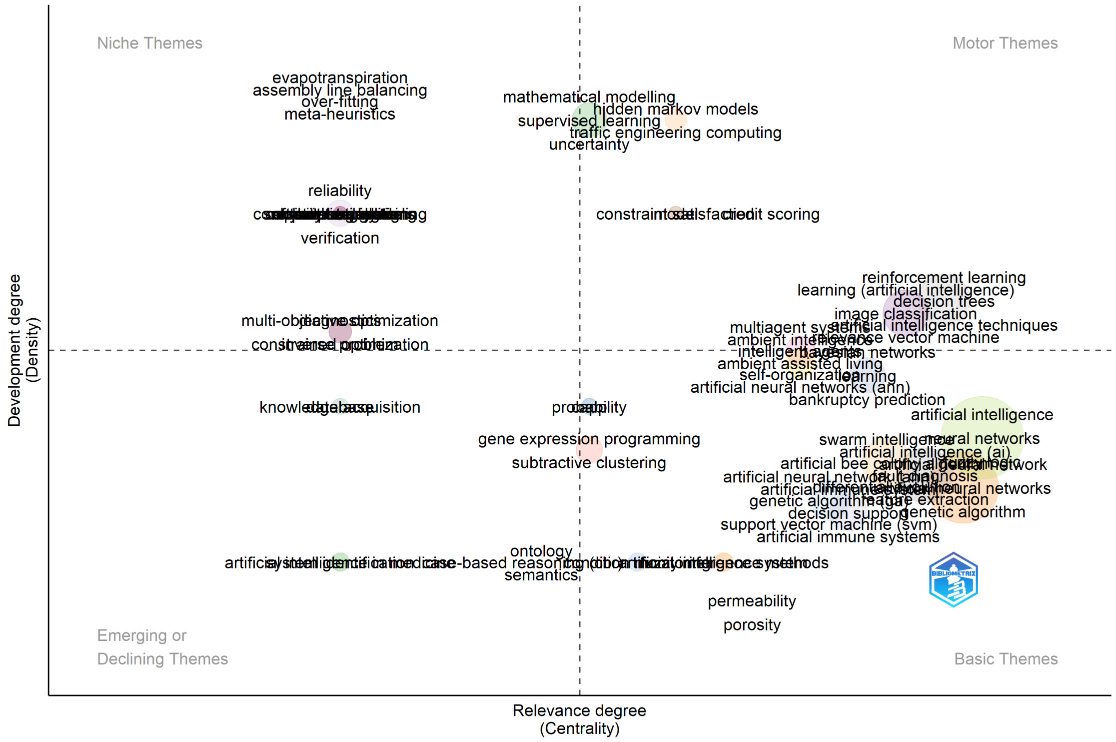

Figure A2. Thematic Evolution of Research in 2005/2015.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 November 2025 doi:10.20944/preprints202511.1933.v1

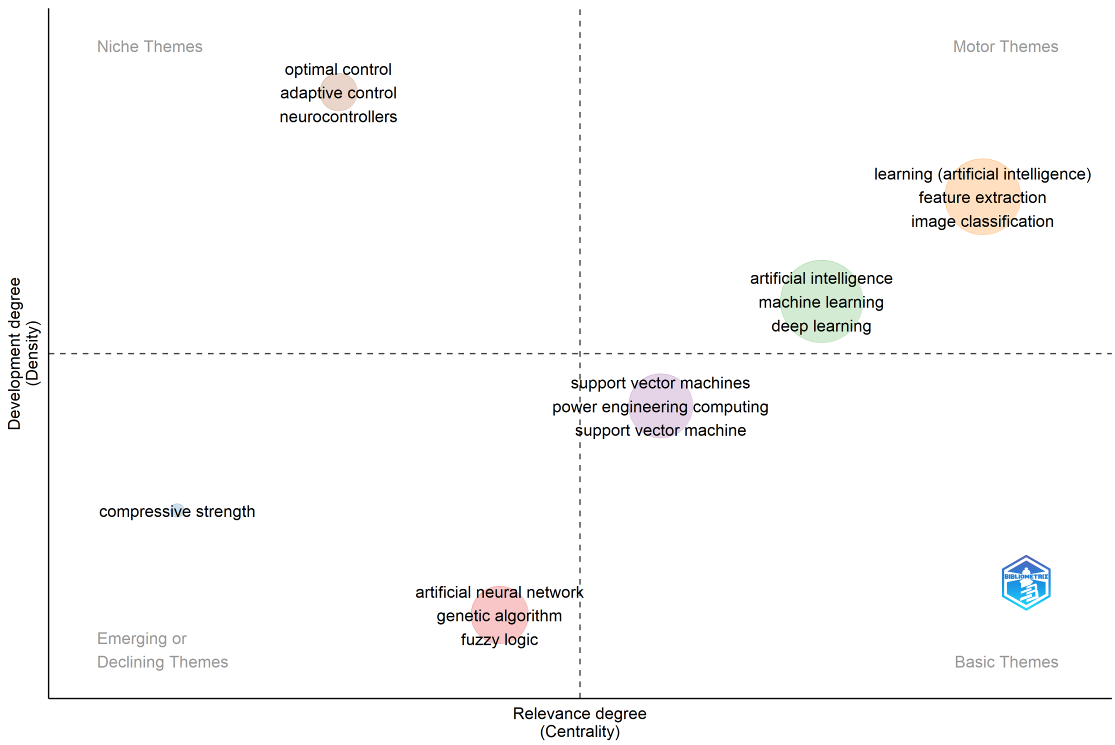

Figure A3. Thematic Evolution of Research in 2016/2020.

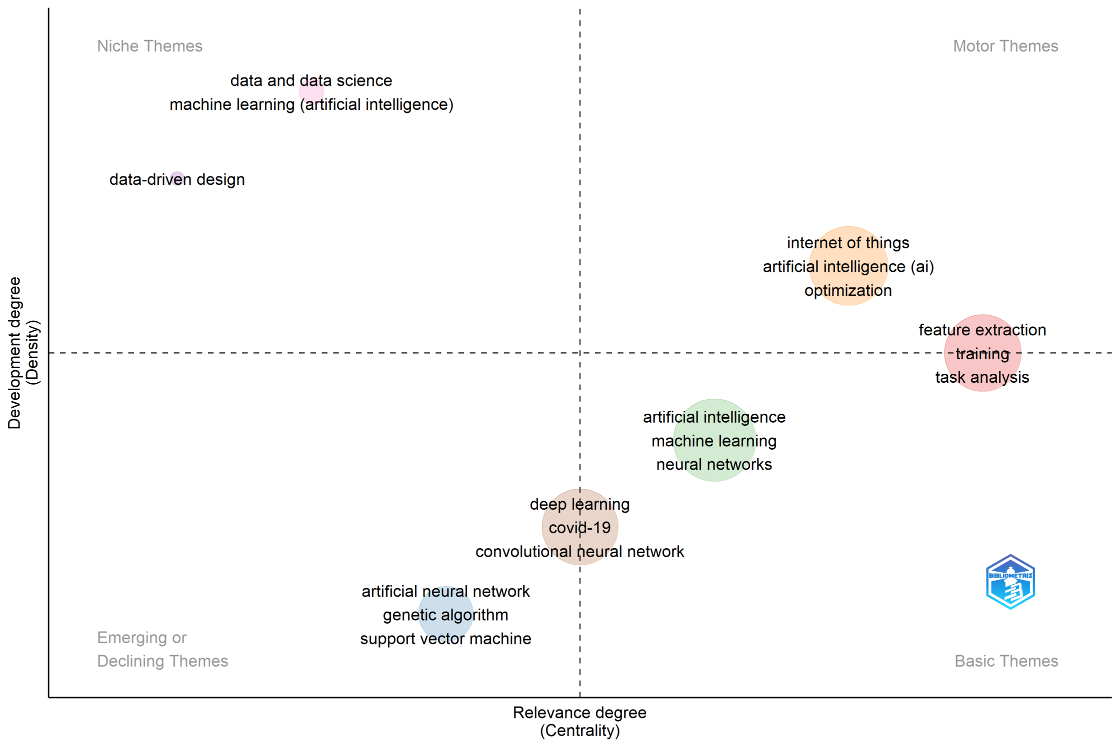

Figure A4. Thematic Evolution of Research in 2021/2022.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 November 2025
doi:10.20944/preprints202511.1933.v1

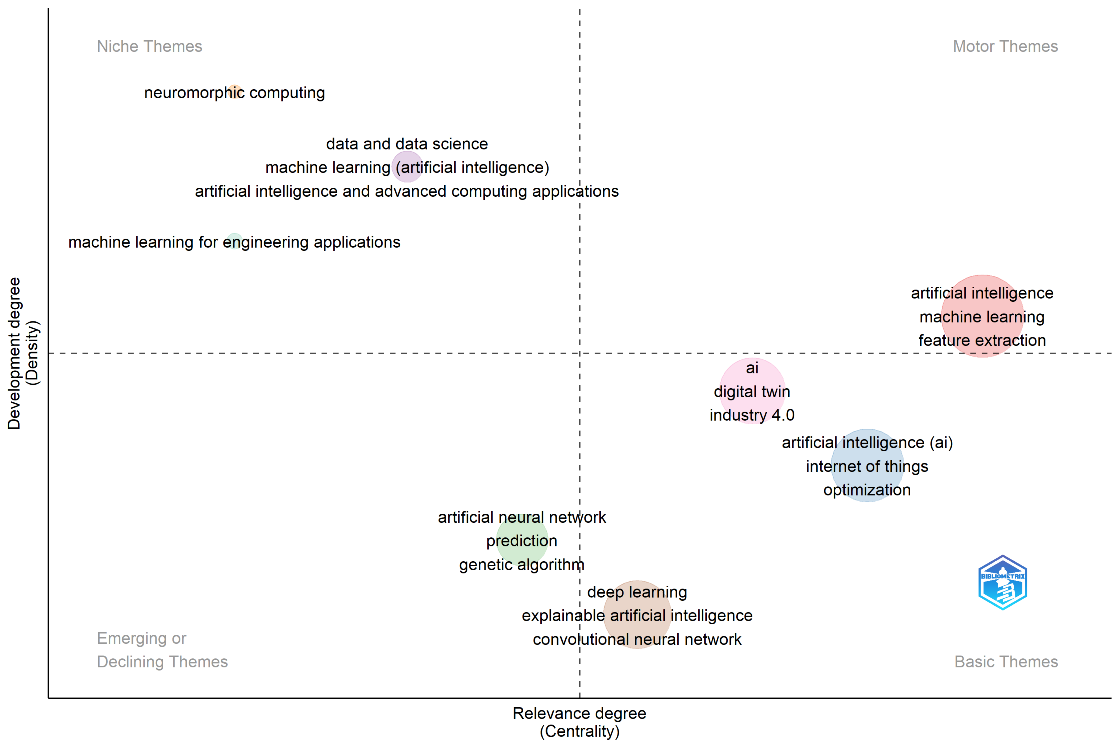

Figure A5. Thematic Evolution of Research in 2023–2024.

References

1. Darko, A. Artificial Intelligence in the Architecture, Engineering, and Construction Industry: A Scientometric Analysis and Future Research Directions. *Automation in Construction* **2019**.
2. Mustak, M.; Salminen, J.; Plé, L.; Wirtz, J. Artificial Intelligence in Marketing: Bibliometric Analysis, Topic Modeling, and Research Agenda. *Journal of Business Research* **2021**.
3. Jiménez-Gómez, C.; et al. Mapping the Literature on the Application of Artificial Intelligence in Library and Information Science. *Library Hi Tech* **2022**.
4. Yaqoob, I. Artificial Intelligence and Machine Learning in Healthcare. *Pain Reports* **2022**.
5. Njei, B.; et al. Artificial Intelligence for Healthcare in Africa: A Scientometric Analysis. *Heliyon* **2023**, *13*, 947–955.
6. Sabanci, A.M. Artificial Intelligence Applications in Sensors: A Scientometric Analysis. *Sensors* **2022**, *22*, 7814.
7. Straub, D. El valor de los estudios cienciométricos: una introducción a un debate sobre el SI como disciplina de referencia. *Journal of the Association for Information Systems* **2006**, *7*, 241–246.
8. Jiménez, M. Apoyo a la investigación: Índice M, 2025. Accessed: 19-Oct-2025.
9. Vardi, M.Y. Conferences vs. journals in computing research. *Communications of the ACM* **2009**, *52*, 5–5.
10. Serenko, A.; Bontis, N. Meta-review of knowledge management and intellectual capital literature: citation impact and research productivity rankings. *Knowledge and Process Management* **2004**, *11*, 185–198.
11. Heilig, L.; Voß, S. A Scientometric Analysis of cloud Computing Literature. *IEEE Transactions on Cloud Computing* **2014**, *2*.
12. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; et al. Generative adversarial networks. *Communications of the ACM* **2020**, *63*, 139–144.
13. Silver, D.; Huang, A.; Maddison, C.J.; et al. Mastering the game of Go with deep neural networks and tree search. *Nature* **2016**, *529*, 484–489.
14. Esteva, A.; Kuprel, B.; Novoa, R.A.; et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **2017**, *542*, 115–118.
15. Dorigo, M.; Birattari, M.; Stützle, T. Ant colony optimization. *IEEE Computational Intelligence Magazine* **2006**, *1*, 28–39.
16. Karaboga, D.; Basturk, B. A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm. *Springer* **2007**, *39*, 459–471.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 November 2025 doi:10.20944/preprints202511.1933.v1

17. Karaboga, D.; Basturk, B. On the performance of artificial bee colony (ABC) algorithm. *Applied Soft Computing* **2008**, *8*, 687–697.
18. Jordan, M.I.; Mitchell, T.M. Machine learning: Trends, perspectives, and prospects. *Science* **2015**, *349*, 255–260.
19. Arrieta, A.B.; Díaz-Rodríguez, N.; Ser, J.D.; Bennetot, A.; et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion* **2020**, *58*, 82–115.
20. Yang, Q.; Liu, Y.; Chen, T.; Tong, Y. Federated Machine Learning: Concept and Applications. *ACM Transactions on Intelligent Systems and Technology (TIST)* **2019**, *10*, 1–19.
21. Silver, D.; Schrittwieser, J.; et al. Mastering the game of Go without human knowledge. *Nature* **2017**, *550*, 354–359.
22. van Griethuysen, J.; Fedorov, A.; Parmar, C.; et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Research* **2017**, *77*, e104–e107.
23. Donthu, N.; Kumar, S.; Mukherjee, D.; Pandey, N.; Lim, W.M. How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research* **2021**, *133*, 285–296.
24. Kirkpatrick, J.; Pascanu, R.; Rabinowitz, N.; Hadsell, R. Overcoming catastrophic forgetting in neural networks. Preprint / pnas.org, 2017.
25. Mannsfeld, S.C.B.; Tee, B.C.K.; Stoltenberg, R.M.; et al. Highly sensitive flexible pressure sensors with microstructured rubber dielectric layers. *Nature Materials* **2010**, *9*, 859–864.
26. Topol, E.J. High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine* **2019**, *25*, 44–56.
27. Musiño, C.M. La ciencia de la información como plataforma para potenciar el estudio de los flujos de la información en las organizaciones. *Revista e-Ciencias de la Información* **2012**, *2*, 1–14.
28. Macías-Chapula, C.A. Papel de la informetría y de la cienciometría y su perspectiva nacional e internacional. *Acimed* **2001**, *9*, 35–41.
29. Adadi, A.; Berrada, M. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access* **2018**, *6*, 52138–52160.
30. Arulkumaran, K.; Deisenroth, M.P.; Brundage, M.; Bharath, A.A. Deep Reinforcement Learning: A Brief Survey. *IEEE Signal Processing Magazine* **2017**, *34*, 26–38.
31. Sze, V.; Chen, Y.H.; Yang, T.J.; Emer, J.S. Efficient Processing of Deep Neural Networks: A Tutorial and Survey. *Proceedings of the IEEE* **2017**, *105*, 2295–2329.
32. Baltrušaitis, T.; Ahuja, C.; Morency, L.P. Multimodal Machine Learning: A Survey and Taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2019**, *41*, 423–443.
33. Abualigah, L.; Diabat, A.; Mirjalili, S.; Elaziz, M.A.; Gandomi, A.H. The Arithmetic Optimization Algorithm. *Computer Methods in Applied Mechanics and Engineering* **2021**, *376*, 113609.
34. Lei, Y.; Yang, B.; Jiang, X.; Jia, F.; Li, N.; Nandi, A.K. Applications of machine learning to machine fault diagnosis: A review and roadmap. *Mechanical Systems and Signal Processing* **2020**, *138*, 106587.
35. Chiang, M.; Zhang, T. Fog and IoT: An Overview of Research Opportunities. *IEEE Internet of Things Journal* **2016**, *3*, 854–864.
36. Barnich, O.; Droogenbroeck, M.V. ViBe: A Universal Background Subtraction Algorithm for Video Sequences. *IEEE Transactions on Image Processing* **2011**, *20*, 1709–1724.
37. Liu, R.; Yang, B.; Zio, E.; Chen, X. Artificial intelligence for fault diagnosis of rotating machinery: A review. *Mechanical Systems and Signal Processing* **2018**, *108*, 33–47.
38. Ozturk, T.; Talo, M.; Yildirim, E.A.; Baloglu, U.B.; Yildirim, O.; Acharya, U.R. Automated detection of COVID-19 cases using deep neural networks with X-ray images. *Computers in Biology and Medicine* **2020**, *121*, 103792.
39. Yu, K.H.; Beam, A.L.; Kohane, I.S. Artificial intelligence in healthcare. *Nature Biomedical Engineering* **2018**, *2*, 719–731.
40. Liakos, K.G.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine Learning in Agriculture: A Review. *Sensors* **2018**, *18*, 2674.
41. Jia, F.; Lei, Y.; Lin, J.; Zhou, X.; Lu, N. Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. *Mechanical Systems and Signal Processing* **2016**, *72–73*, 303–315.
42. Zhao, Z.; Chen, W.; Wu, X.; Chen, P.C.Y.; Liu, J. LSTM network: a deep learning approach for short-term traffic forecast. *IET Research* **2017**, *11*, 68–75.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 November 2025 doi:10.20944/preprints202511.1933.v1

43. M, A.; C., C. Bibliometrix: an R-tool for comprehensive science mapping analysis. *J Informetr* **2017**, p. 959–975.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.