Mapping National Artificial Intelligence Research Authority

A Scientometric Analysis of AI Subdisciplines Using OpenAlex (2020–2025)

Julius Pfundstein

Faculty of Mathematics and Computer Science

University of Leipzig

julius.pfundstein@studserv.uni-leipzig.de

UNIVERSITÄT
LEIPZIG

Oktober 2025

Acknowledgments

I would like to express my deepest gratitude to my supervisor,
Dr. Thomas Efer,
for his invaluable guidance, continuous support, and insightful feedback throughout the course of my master's thesis.

I also wish to sincerely thank the OpenAlex team for generously providing access to their API, which enabled the extensive data collection and analysis crucial to this research.

Faculty of Mathematics and Informatics
University of Leipzig
Master of Science in Digital Humanities

Abstract

Artificial intelligence (AI) has rapidly emerged as a foundational general-purpose technology, profoundly reshaping productivity across multiple sectors including manufacturing, services, healthcare, finance, and defense. Beyond optimizing discrete tasks, AI increasingly reconfigures the institutional and geopolitical frameworks through which knowledge, value, and power are produced and contested. This raises critical questions about the distribution of technological authority within the global AI research ecosystem.

This study addresses two central research questions: (1) Which subdisciplines compose the contemporary field of AI research? and (2) How are countries and regions positioned within these subfields in terms of research capacity and influence? Utilizing a comprehensive dataset of over three million AI-related publications from OpenAlex, we implement hierarchical topic modeling on titles and abstracts to delineate distinct AI subdisciplines. We then apply citation network analysis to measure the relative centrality and scholarly impact of countries within each subfield.

By integrating thematic and network-based bibliometric methods, this work offers a detailed, subdiscipline-specific mapping of the global AI research landscape. The findings illuminate the complex structure of AI as a multifaceted domain and reveal the geographic patterns of research leadership, providing a critical empirical foundation for understanding the evolving loci of AI-driven power and influence.

Keywords: Artificial Intelligence, Bibliometrics, Scientometrics, Global AI Research, Country Ranking

CONTENTS

1. INTRODUCTION

Artificial intelligence has emerged as a general-purpose technology, woven deeply into the fabric of contemporary productivity. Much like the electrification of industry once redefined the foundations of economic growth, AI now underpins advances across manufacturing [1], services [2], healthcare [3], finance [4], and defense [5] [6]. Its pervasive integration means that AI does not merely optimize isolated tasks but increasingly conditions the structures through which knowledge, value, and strategic advantage are generated and contested. Technologies of such systemic importance inevitably transcend the technical; they become instruments through which power is organized, exercised, and shifted within and between states.

For political science, this raises an enduring yet newly urgent inquiry: *where does the power embedded in this technological capacity reside, and how is it distributed across the globe?*

Addressing such questions empirically calls for tools capable of systematically tracing patterns of knowledge production, diffusion, and influence across large-scale scientific activity. Scientometrics, as the quantitative study of science and scholarly communication, provides a robust framework for capturing these dynamics by analyzing publication data, citation networks, and thematic structures within research fields.

A review of the existing literature reveals substantial efforts to map the global landscape of artificial intelligence research. This body of work has produced valuable insights into national capacities, collaboration patterns, and knowledge flows. However, much of this research remains limited in at least one of five critical respects:

1. Analyses relying on outdated data fail to capture recent developments and dynamics within the field. [7]
2. Studies with narrow application focus—such as AI in life sciences [8] or innovation [9]—neglect the comprehensive disciplinary structure.
3. Conceptualizing AI as a unified domain obscures the heterogeneity of subfields characterized by distinct methodologies and strategic priorities. [10, 11, 12, 13, 14]
4. Scientometric analyses that prioritize authorship-level metrics lack systematic evaluation of national and regional research positioning.
5. National AI capacity assessments typically exclude multi-dimensional benchmarking frameworks beyond scientometric indicators. [15]

Taken together, these limitations illustrate why existing work remains insufficient for capturing the full scope and complexity of contemporary artificial intelligence research. Analyses constrained by outdated data, narrow disciplinary scope, or overly aggregated concepts risk masking the field's internal differentiation and its shifting centers of expertise. Likewise, a sole focus on authorship-level patterns or on single indicators overlooks the multiple scales and dimensions through which technological capacity is organized and expressed. Addressing these gaps requires an approach that combines fine-grained topical resolution with systematic, comparative assessment at the country level.

To address this limitation, two guiding research questions structure the present analysis:

RQ1 *Which subdisciplines structure the contemporary field of artificial intelligence research?*

RQ2 *How are countries and regions positioned within these subdisciplines in terms of their relative research capacity and influence?*

These questions serve as the foundation for the empirical strategy and frame the subsequent analysis.

To address these research questions, this study utilizes a comprehensive dataset comprising over three million artificial intelligence research publications sourced from OpenAlex. [16] A hierarchical topic modeling approach is applied to the titles and abstracts to identify and characterize the distinct subdisciplines within AI. Subsequently, a citation network analysis quantifies the relative centrality and influence of countries and regions within these identified subfields. This combined methodology enables a nuanced mapping of the global AI research landscape, providing empirical insight into both the internal structure of AI as a research domain and the geographic distribution of scholarly leadership.

2. RESEARCH REVIEW

The global distribution of research capacities in artificial intelligence has attracted increasing scholarly and policy attention in recent years. Numerous studies have examined national investments, scientific output, and patterns of collaboration in AI-related fields, reflecting concerns about technological leadership, economic competitiveness, and strategic autonomy [**REF**]. Comparative assessments often rely on publication counts, patent registrations, or research funding indicators to measure countries' positions within the global AI landscape [**REF**].

This body of work has provided valuable insights into the dynamics of knowledge production, the role of leading institutions, and the emergence of new centers of AI expertise [**REF**]. However, a common limitation of many large-scale comparative studies is their tendency to treat artificial intelligence as a single, unified domain. Such an approach

risks masking substantial heterogeneity within AI, which spans diverse technical domains, application areas, and research communities. Recent contributions have begun to recognize this internal diversity, for example by mapping thematic clusters or tracing the evolution of specific subfields such as deep learning or natural language processing [**REF**]. Nevertheless, systematic and scalable analyses that combine topic modeling with cross-national comparison remain limited.

In addition to thematic differentiation, the question of scientific influence and leadership is often approached through simple output metrics, such as publication counts or citation aggregates at the national level. While these measures capture research volume, they can obscure structural aspects of influence within scientific networks. Network-based indicators — such as citation centrality — offer a more nuanced perspective by highlighting how ideas, methods, and findings diffuse and concentrate across communities [**REF**].

Against this backdrop, the present study aims to contribute to the literature by systematically disaggregating the AI research domain into empirically derived subdisciplines and situating countries within these subfields based on their structural position in global citation networks. This approach complements existing work by combining large-scale topic modeling and network analysis to capture both the thematic complexity of AI and the relational dynamics that underpin scientific leadership.

3. DATA

This chapter outlines the systematic approach undertaken to assemble a robust and representative dataset of artificial intelligence research publications. It begins by discussing the selection criteria for the underlying scientific knowledge graph, followed by the construction of a comprehensive search query based on an extensive review of recent AI survey literature. Subsequent sections detail the data retrieval process, post-processing steps including duplicate removal and metadata validation, and the preparation of the final corpus for analysis. This structured approach ensures that the dataset aligns closely with the study's objectives and provides a reliable foundation for the applied topic modeling and citation network methods.

3.1. SKG-COMPARISON

Analyses of technological capacity and leadership often rely on empirical indicators such as economic output, patent filings, or scientific publications. While economic and patent data capture dimensions of technological application and commercialization, they offer only limited insight into the underlying processes of knowledge generation and scholarly exchange that drive emerging research frontiers. In the context of artificial intelligence — a domain where conceptual advances and methodological innovations are tightly interwoven with practical applications — peer-reviewed research outputs provide a particularly suitable proxy. Scientific publications document not only the scale of research activity but also its thematic orientation, methodological diversity, and the collaborative structures through which knowledge is produced and diffused.

To examine these dynamics systematically, this study employs OpenAlex as the principal data source. OpenAlex belongs to the class of Scientific Knowledge Graphs (SKGs), which integrate scholarly metadata, citation relationships, institutional affiliations, and research topics into structured, interlinked databases. Prominent SKGs include Web of Science, Scopus, and Microsoft Academic Graph (MAG). Compared to proprietary platforms, OpenAlex offers broader disciplinary coverage, strong representation of recent and interdisciplinary works, and open programmatic access that supports scalable, reproducible analysis.

A comparative overview of leading SKGs is provided in Table 1, highlighting the features that make OpenAlex particularly well suited to address the present research questions concerning thematic differentiation and network-based measures of scholarly influence.

Among the available Scientific Knowledge Graphs, OpenAlex offers specific advantages that align closely with the analytical goals of this study. Unlike many proprietary databases that prioritize indexing high-impact journals and rely on selective curation, OpenAlex adopts an inclusive indexing strategy that reduces coverage bias toward established or elite institutions. This broader scope enhances the visibility of contributions from emerging research contexts, including non-OECD countries and institutions outside traditional centers of scientific production.

Moreover, OpenAlex provides the strongest representation of non-English language outputs among major SKGs, which is critical for capturing regional and national specializations that may be underrepresented in predominantly Anglophone collections. Its comprehensive metadata — including consistent DOI assignment, author identifiers, abstracts, referenced works, and citation counts — ensures the completeness and comparability required for robust topic modeling and network analysis.

Finally, OpenAlex's open-access licensing and unrestricted API facilitate scalable data collection and reproducibility, both of which are essential for large-scale scientometric studies. Taken together, these characteristics position OpenAlex as the most suitable source for examining the internal structure of artificial intelligence research and the relative standing of countries within this dynamically evolving field.

Table 1: Comparison of Major Scientific Knowledge Graphs

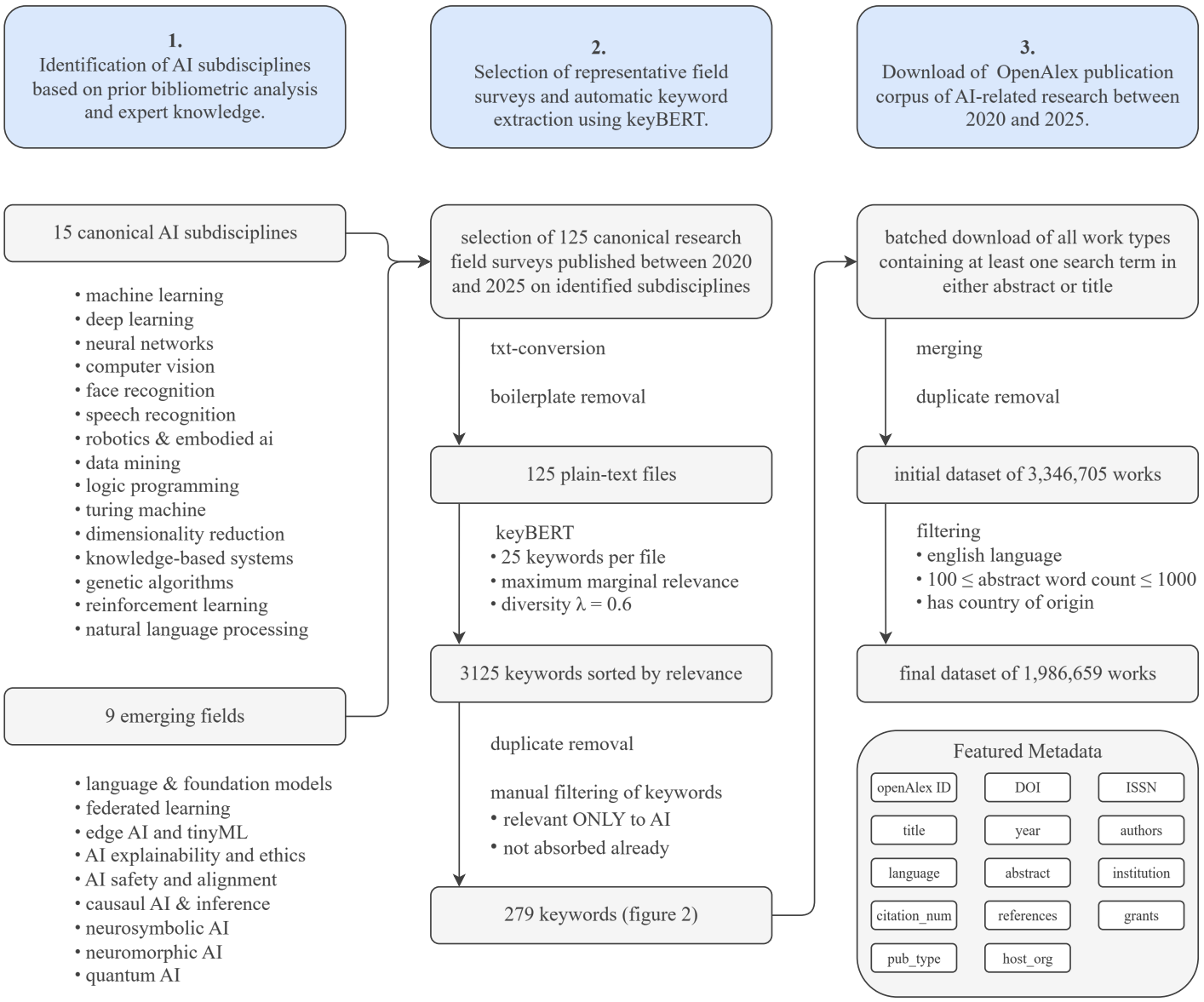
Metric	Web of Science	Scopus	Dimensions	OpenAlex
Basic Coverage				
Total documents (millions)	90	94	134	248
Metadata Completeness				
DOI coverage	98%	97%	95%	99%
Author ORCID linkage	Partial	Yes	Yes	Yes
Abstract availability	High	High	High	High
Referenced works coverage	High	High	High	High
Citation count availability	High	High	High	High
Geographic Coverage				
English-language dominance	High	High	Moderate	Low
Non-English abstracts	15%	25%	30%	40%
Developing country coverage	Moderate	Good	Excellent	Good
Access & Features				
API access	Restricted	Restricted	Full	Full
Open access	No	No	Partial	Yes
Cost model	Subscription	Subscription	Freemium	Free
Text mining allowed	Limited	Limited	Yes	Yes

Note: The data presented in this table reflect the state of the respective scientific knowledge graph platforms as closely as possible to the time of dataset compilation in June 2025. Given the dynamic and continuously evolving nature of these resources, values may be subject to temporal variation and periodic updates.

It must be acknowledged, however, that all Scientific Knowledge Graphs, including OpenAlex, reflect inherent biases related to language, regional publishing practices, and coverage policies. Recent comparative assessments (e.g., [**REE**]) have highlighted systematic differences across SKGs that can shape analytical outcomes. Recognizing these limitations is essential for interpreting the results presented in this study.

3.2. EXTRACTION PIPELINE

Figure 4.1: Dataset Construction Pipeline Diagram



EXTRACTION PIPELINE

The selection of publications is a crucial determinant of this study's analytical validity. As Baruffaldi et al. propose, building a comprehensive query for artificial intelligence research requires a structured keyword set; their 2020 study derived a list of 193 AI-related terms from AI-labeled publications from Scopus. [17] However, given the rapid evolution of the field, such a list must be continuously revised to reflect emerging subdisciplines and novel applications. To address this, a similar but updated approach was undertaken here. The extraction process began with a framework established by Gargiulo et al. [18], who applied topic modeling to delineate 15 principal AI subfields that collectively define the core technological landscape. These subfields form the initial structure for constructing a refined and thematically precise search strategy, which is detailed in the following sections.

(1) Neural networks, (2) deep learning, (3) machine learning, (4) reinforcement learning, (5) natural language processing, (6) computer vision, (7) face recognition, (8) speech recognition, (9) robotics, (10) data mining, (11) logic programming, (12) Turing machines, (13) dimensionality reduction, (14) expert systems, (15) and genetic algorithms.

Building on this taxonomy, an intensive literature review identified an additional nine emergent subdisciplines that have gained significant attention post-2022, reflecting the rapid evolution of AI research. These include, among others, (16) language models and foundation models, (17) federated learning, (18) edge AI, (19) AI explainability and ethics, (20) AI safety and alignment, (21) causal AI, (22) neurosymbolic AI, (23) neuromorphic AI, (24) and quantum AI.

For each identified subdiscipline, a corpus of survey papers was assembled, restricted to publications from 2019 onward to ensure topical relevance. Survey papers were selected because they offer comprehensive overviews of their respective fields, cite a wide range of seminal works, and avoid overly specialized or niche vocabulary that might exclude broader field indicators. In total, 125 survey papers were collected.

From these survey papers, keywords were extracted using a modified version of the KeyBERT algorithm. This approach yielded an initial set of 3,175 candidate keywords, which underwent manual refinement to isolate the 279 terms most exclusively relevant to artificial intelligence research. This refined keyword list served as the basis for constructing the search query applied to the OpenAlex database.

A simplistic keyword query such as "Artificial Intelligence" was initially tested but proved insufficient. An exact keyword match on titles and abstracts returned only approximately 570,714 works, many of which were tangential to the technological aspects of AI. Analysis of OpenAlex-provided concept distributions in a random sample of 10,000 works revealed that a substantial fraction pertained to non-technical domains, including economics (3.0%), political science (2.3%), law (2.2%), sociology (1.0%), and philosophy (0.9%). While these areas offer valuable perspectives, the focus of this study on technological capacities necessitated minimizing their representation to enhance precision without sacrificing recall.

To further address this challenge, the final keyword set deliberately excludes certain broad or cross-disciplinary concepts—such as *multi-agent systems*, *natural language processing*, *genetic algorithms*, and *expert systems*—which, despite their centrality within AI, also have substantial relevance in other domains that could introduce thematic noise and inflate the dataset with unrelated documents. Likewise, generic umbrella terms like *artificial intelligence* itself were omitted for the same reason: they frequently appear in social science publications that seek to contextualize the phenomenon rather than contribute to its technical development. Acronyms such as *GAN*, *CNN*, *RLHF*, *GNN*, and *SNN* were also excluded, as they often have multiple domain-specific meanings and can generate ambiguous matches, particularly given that the search is case-insensitive and abbreviations can appear embedded within unrelated words (for instance, *RAG* in *fragmentation*). Even the standalone term *AI* was omitted except when paired with additional descriptors that unambiguously anchor it within computer science.

Conversely, some keywords were retained in the final list even though they cannot be exclusively attributed to AI research. For instance, methods such as *backpropagation* or *random forest* are also widely used in broader statistical learning and other related fields. However, given their strong association with AI-driven machine learning tasks, it was assumed that their inclusion would primarily yield AI-relevant publications, thus justifying their retention. By contrast, terms for which a clear or probable secondary meaning exists—such as *BERT* or *PaLM*—were excluded to minimize ambiguity and unintended retrieval.

This cautious curation ensures that the resulting corpus remains as technically focused as possible while minimizing false positives. Consequently, the comprehensive keyword list was constructed with the dual objective of maximizing recall for AI-relevant publications and minimizing the inclusion of unrelated works. The list was implemented in a Boolean `OR` query to retrieve a broad yet targeted dataset for subsequent analysis.

A visualization of the number of retrieved results per individual search term is provided in Figure 4.1. The fact

that the total number of retrieved documents falls short of the sum of the individual term results is explained by the overlap inherent in the query design: multiple search terms frequently co-occur in the title and abstract of a single document.

3.2.1 Postprocessing

In the postprocessing stage, duplicate records were identified and removed using the unique internal identifiers provided by OpenAlex. This step ensured that each scholarly work was represented only once in the dataset, eliminating redundancy and potential skewing of results.

Subsequently, the completeness of the metadata fields was systematically evaluated. Given that individual works can contribute to multiple methodological categories, filtering out records solely based on missing fields such as abstracts or author information was deliberately avoided. This approach minimizes the risk of introducing unknown biases that could arise from the selective exclusion of partially incomplete records.

The resulting curated dataset was stored in a structured SQL database comprising 3,345,732 works. For each work, the following metadata fields were retained and are available for analysis:

- **Authors:** including names, ORCID identifiers, and affiliated institutions
- **Referenced Works:** bibliographic references cited by the work
- **Publisher:** name of the publishing entity
- **Publishing Year:** year of publication
- **Citation Count:** number of citations received
- **Funder Information:** details on funding bodies and funding amounts
- **Journal Type:** classification of the publication venue
- **Open Access Status:** binary indicator specifying whether the work is openly accessible

From the authors' institutional affiliations, a *country of origin* variable was derived. This variable captures the geographic distribution of all contributing author institutions per work, represented as a list of tuples containing country codes and their respective shares of total author institutions. For example:

```
[("CN": 0.5), ("US": 0.25), ("DE": 0.25)]
```

indicates that 50% of the authors' institutions are located in China, 25% in the United States, and 25% in Germany. This representation facilitates nuanced analyses of international collaboration patterns and geographic contributions at the institutional level.

4. METHODOLOGY

The methodological framework of this study is divided into two complementary domains, directly aligned with the structure of the research questions.

First, a hierarchical topic modeling approach is applied to the titles and abstracts of all works contained in the curated dataset. The objective of this stage is to algorithmically identify latent subdisciplines within the broader field of artificial intelligence research. By leveraging topic modeling, the study uncovers coherent thematic clusters that reveal the evolving structure of AI subfields as they emerge from textual content.

Second, the study quantifies the relative research authority of individual countries within the subdisciplines identified in the first stage. This is accomplished through a citation-network analysis that computes centrality measures for the most recent research outputs. By constructing directed citation graphs and analyzing their topological properties, the study derives metrics of scholarly influence and connectivity, thereby providing an empirical basis for assessing the international distribution of research leadership in AI.

The following subsections describe these two methodological components in detail.

4.1. SUBDISCIPLINE RETRIEVAL

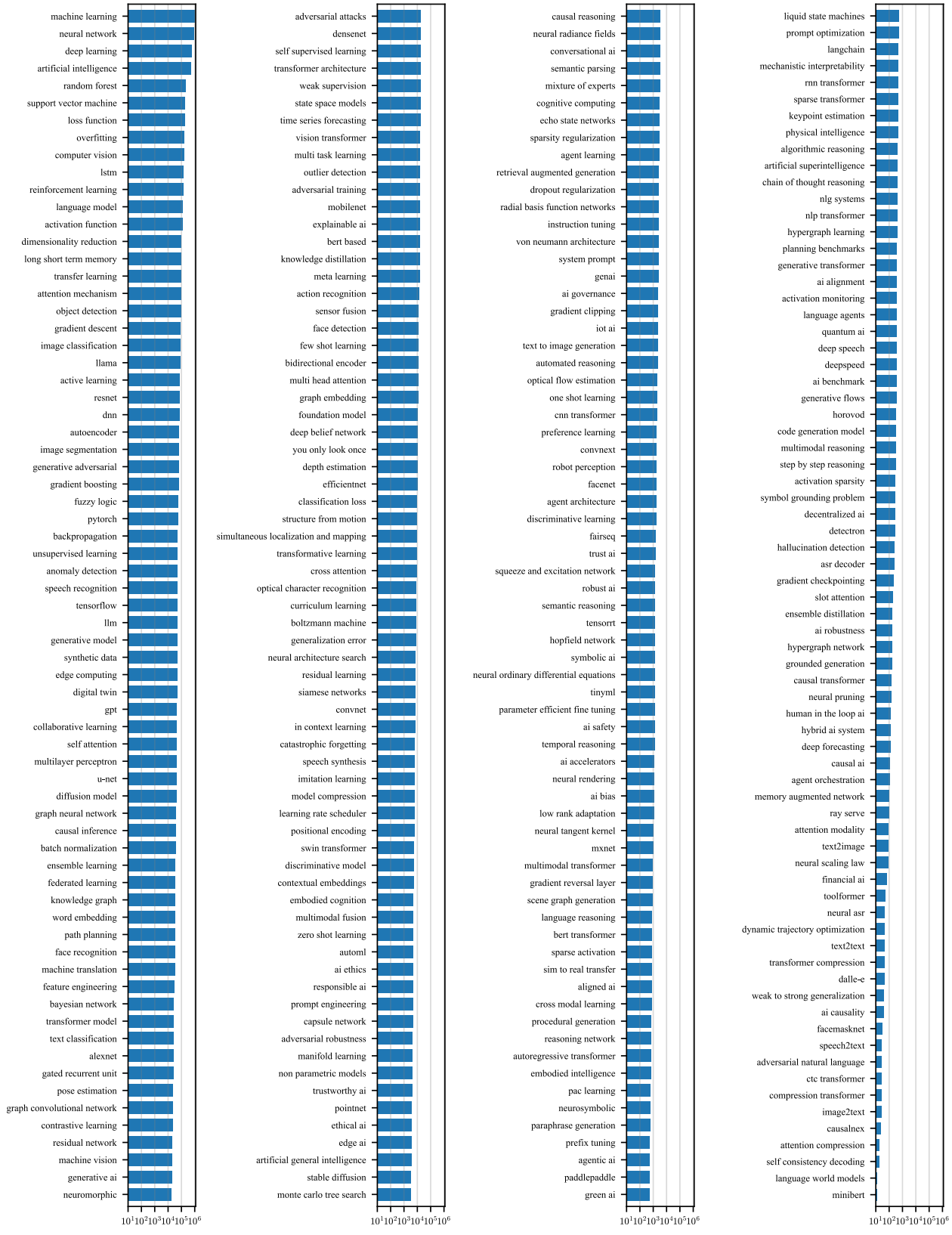
The identification of AI subdisciplines is operationalized through a multi-stage hierarchical topic modeling pipeline, combining feature extraction, dimensionality reduction, clustering, and semantic labeling.

To begin, the textual corpus for each work is constructed by appending the title twice to the abstract. This weighting reflects the well-founded assumption that the title is a concise yet highly informative representation of the work's core contribution, thereby increasing its influence in the embedding space.

For the text vectorization, the `scikit-learn feature_extraction` module is employed to transform the preprocessed corpus into high-dimensional embeddings. These embeddings form the basis for subsequent clustering.

A key step in this pipeline is the determination of the number of clusters k. While there exist data-driven heuristics for estimating an "optimal" k (see, e.g., [19]), the notion that k is inherent to the data alone is often methodologically fragile and prone to post-hoc rationalization. Conse-

Figure 1: Logarithmic bar graph of retrieval count per search term



quently, this study adopts a transparent and reproducible approach, treating k explicitly as a tunable hyperparameter subordinate to the research design.

The principal criterion for selecting k is practical interpretability and thematic coherence. To this end, UMAP projections are generated to visualize the emergent thematic landscape and to assess the integrity of meaningful clusters. The final k is chosen to ensure that naturally cohesive thematic areas—such as *Medical AI* and *Education AI*—remain intact without unnecessary fragmentation (e.g., separating *Cancer Diagnosis* and *Medical Image Segmentation*), while ensuring that highly specific research fronts such as *Edge AI* or *Quantum AI* do not dissolve into overly diffuse catch-all clusters.

After clustering, each cluster is assigned an interpretable subdiscipline label. This is accomplished by extracting representative keywords that balance two competing criteria: *Distinctiveness*—terms that differentiate the cluster clearly from others—and *Representativeness*—terms that accurately describe the broader thematic scope of the cluster. This dual criterion ensures that labels do not become overly narrow (overemphasizing idiosyncratic terms) nor overly generic (diluting the thematic specificity), but rather capture the defining characteristics of each subdiscipline in a linguistically meaningful and reproducible manner.

Figure 2: Abstract Token Length Distribution

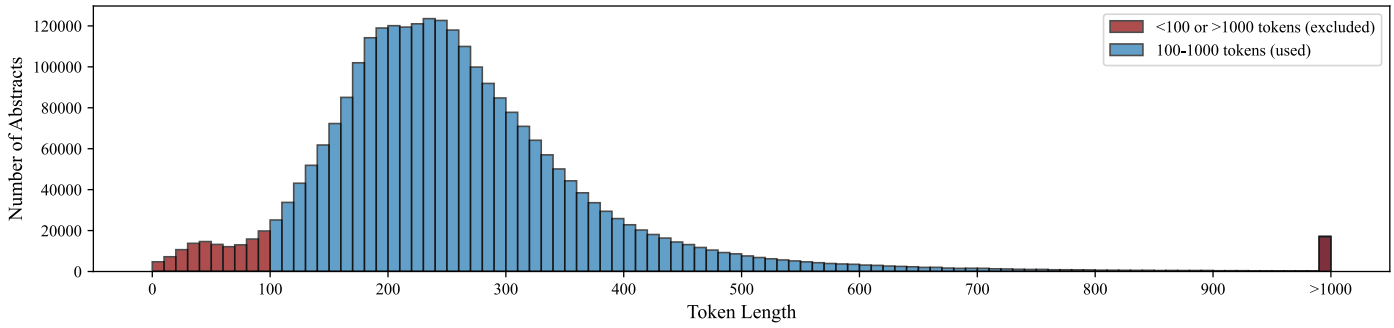
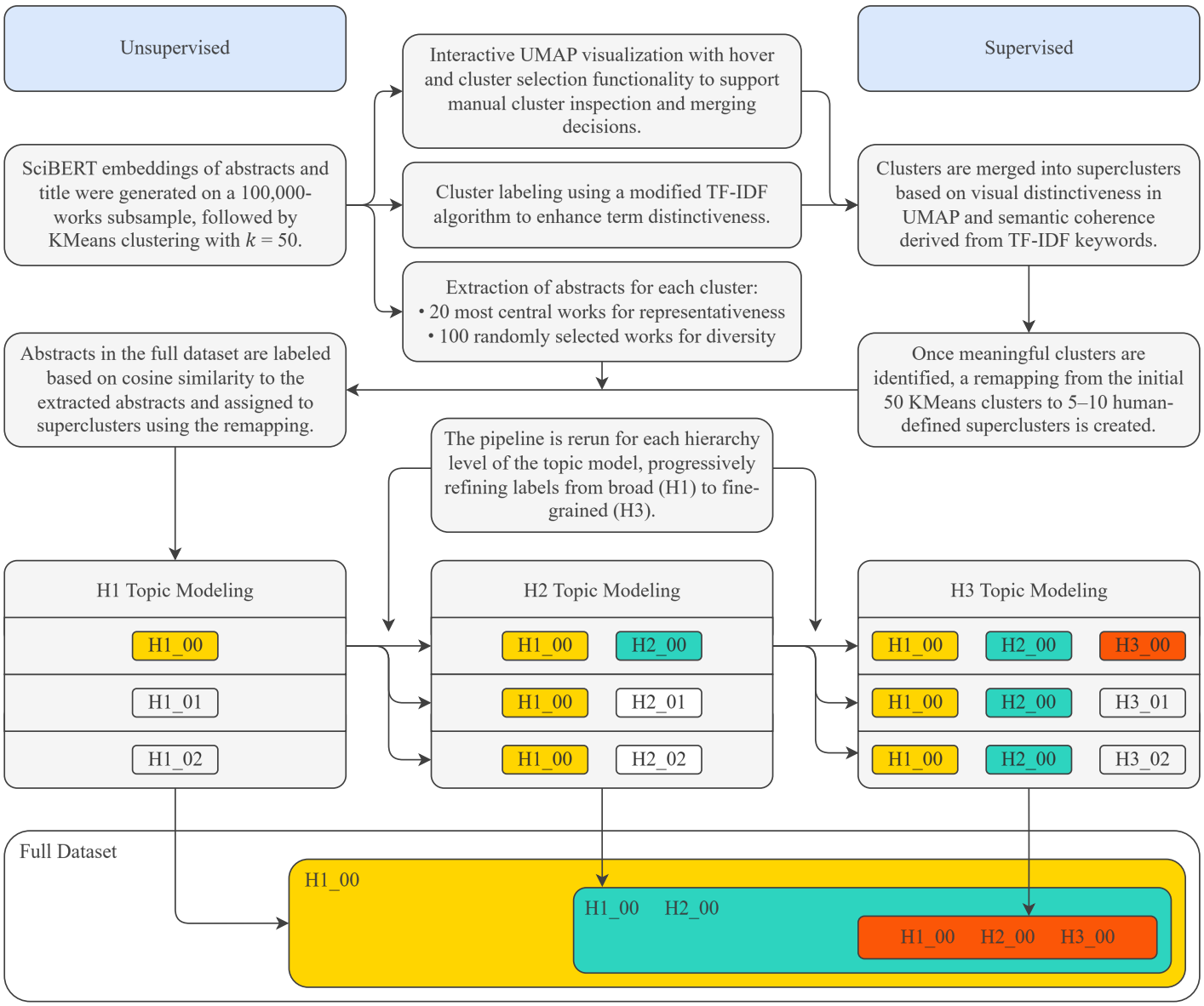


Figure 4.1: Topic Modeling Pipeline Diagram



4.2. QUANTIFICATION OF NATIONAL RESEARCH AUTHORITY

To quantify the relative research authority of countries within each identified subdiscipline, a composite impact score was computed for every work in the dataset. This score integrates both traditional citation impact and structural network centrality, weighted to reflect the influence of each publication within its thematic context.

For each work p in subdiscipline s, the normalized impact score is defined as follows:

$$\text{Impact}_{p,s} = \alpha \times \text{NormCitations}_{p,s} + \beta \times \text{NormCentrality}_{p,s} \tag{1}$$

Here, α and β are adjustable coefficients that determine the relative contribution of citation counts and network-based authority measures to the overall impact metric.

The normalized citation component is calculated by scaling each paper's raw citation count relative to the mean citation count of all works published in the same year and subdiscipline:

$$\text{NormCitations}_{p,s} = \frac{\text{Citations}_p}{\text{MeanCitations}_{\text{year},s}} \tag{2}$$

This normalization corrects for temporal citation accumulation effects and ensures comparability across publication years.

The network-based component, $\text{NormCentrality}_{p,s}$, is derived from the structural position of each work within the directed citation network. This measure can be operationalized using node-level metrics such as PageRank, which captures both direct and indirect citation flows. The centrality score is then normalized by the maximum or mean value within the respective subdiscipline:

$$\text{NormCentrality}_{p,s} = \frac{\text{Centrality}_p}{\max(\text{Centrality}_s) \quad \text{or} \quad \text{MeanCentrality}_s} \tag{3}$$

To attribute this impact to countries, each paper's composite impact score is proportionally distributed according to the institutional affiliations of its authors. The contribution of each country c is calculated by multiplying the paper's impact score with the share of authors' institutions located in that country, as captured by the country distribution vector described in the postprocessing stage.

This procedure produces a subdiscipline-specific authority profile for each country, combining bibliometric performance and network embeddedness with granular affiliation data to yield a robust measure of national research influence within the global AI landsca

5. RESULTS

5.1. HIERARCHICAL TOPIC MODELING

5.2. GEOMAPPING OF NATIONAL AI-AUTHORITY

6. CONCLUSION

7. THINGS I WANT TO ADD

-scientometric usually not performed on super recent papers, but here its okay because the dataset is very big AND discipline very focused on newest developments.

-96383209 edges and 22150971 nodes in network of all works

-there is no raw data, patents, publications and every other indicator already for becoming an indicator is prone to distrortion. china is often overproportionate publishing and patenting while bussiness lacks behind.

-caveat section: at the same time us-papers are more cited just for pathdependancy of scientific cooperation not content only.

-at the same time papers have some benefits over other metrics: no inflation-parity calculus,

-more alarmist introduction: reference ai 2027 paper

-speculation of structural differences: us proprietary, dense models, cn open source sparce and mixture of experts (training on benchmarks and training on api-usecases as one lives from shareholders and the other from market early on)

-filtering method for abstracts: language en, abstract non NULL, country of origin non NULL, abstract more than 99 and less than 1001 tokens.

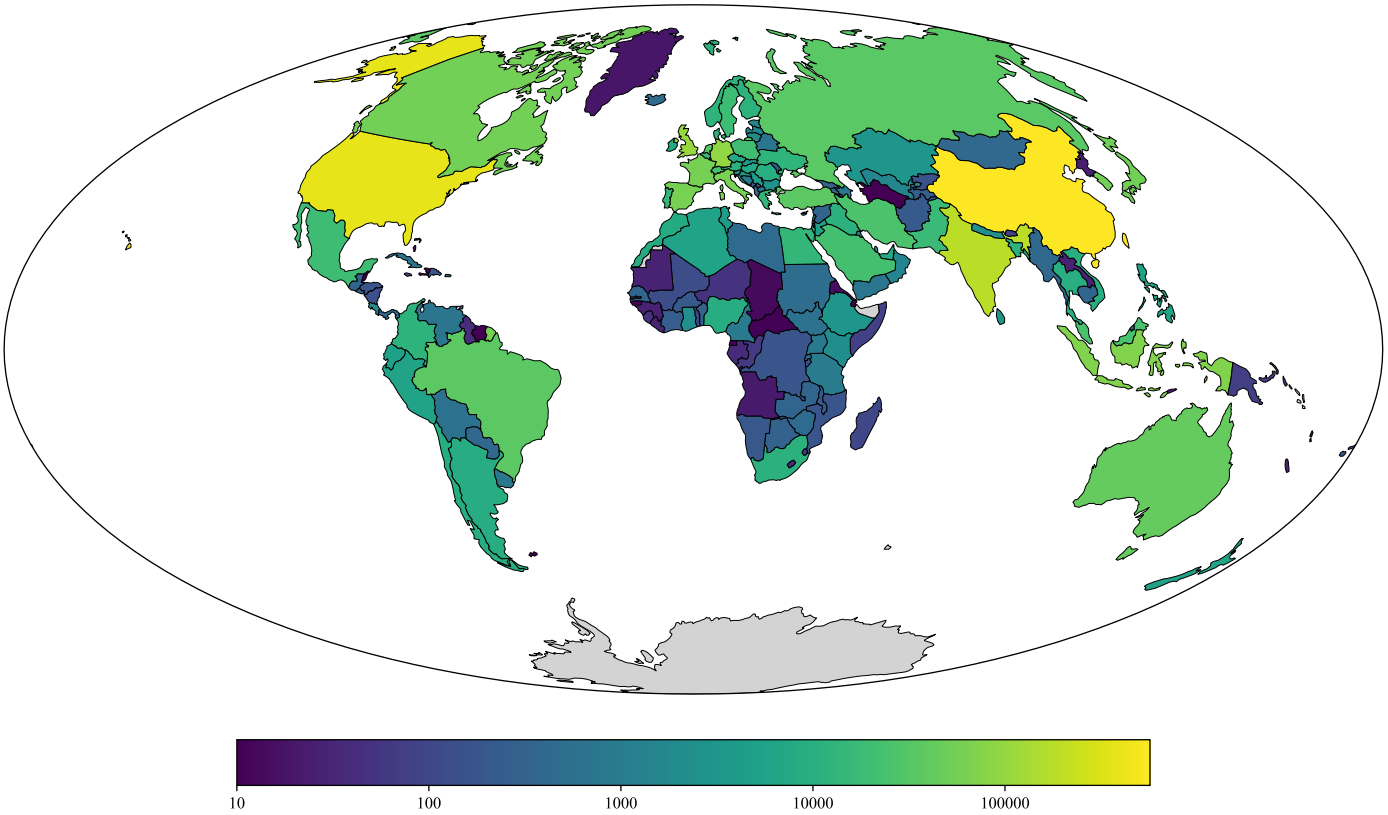
-the logic behind using both most central and random abstracts in a cluster summary is based on complementarity between representativeness and diversity:

-arbitrary in nature: medial image segmentaiton is BOTH - medical and patter recognition - expert have to decide or choose aesthetic criteria.

-human in the loop-curation: Enables top-down supervision while retaining bottom-up discovery.

- clustes are not mutually-exclusive and jointly-exhaustive, but rather best possible descriptor of cluster (decided by UMAP and tfidf terms).

Figure 3: Logarithmic distribution of AI-publications by country



- there are domain, field, subfield and topic labels (topic comes first rest is based on that) - but they dont work really well, mainly because model worked on all papers and could be optimized with field specific work (huamn in the loop). also there is a fwci score but fwci doesnt really work for super young paper.

- idea for dominance measurement: how much more citations than expected (median)?

- cluster names were created by looking at the most central abstracts and the tfidf word supplied in the interactive chart.

- caveat: the error stacks up with each hierarchy, even in case of 90 percent percision one would get onle 72.9 percent in a h3 cluster

- words of abstract contain markers for both method and domain - further research - split up into either of those and try to minimise noise by the other.

-quality control maybe testing against the topics from openalex.

- one word can only capture so much - therefore cluster names are often only neighbours to similar issues (example: there are only about 4k paper that are considered tourism - too little to justify its own category, therefore its in the urban development cluster which has the least cosine distance but doesnt reflect that fact).

- also names can appear twice: geospatial mapping is both a method and an application

- supervision is done on unfiltered umap - no artificial distinctivness enhancement

- density based cluster summarization

REFERENCES

[1] Sung Wook Kim et al. "Recent Advances of Artificial Intelligence in Manufacturing Industrial Sectors: A Review". In: *International Journal of Precision Engineering and Manufacturing* 23 (2022), pp. 111–129. DOI: 10.1007/s12541-021-00506-9. URL: https://doi.org/10.1007/s12541-021-00506-9.

[2] Ming-Hui Huang and Roland T. Rust. "Artificial Intelligence in Service". In: *Journal of Service Research* 21.2 (2018), pp. 155–172. DOI: 10.1177/1094670517752459. URL: https://doi.org/10.1177/1094670517752459.

[3] Ahmed Al Kuwaiti et al. "A Review of the Role of Artificial Intelligence in Healthcare". In: *Journal of Personalized Medicine* 13.6 (2023), p. 951. DOI: 10.3390/jpm13060951. URL: https://doi.org/10.3390/jpm13060951.

[4] Salman Bahoo et al. "Artificial intelligence in Finance: a comprehensive review through bibliometric and content analysis". In: *SN Business & Economics* 4 (2024), p. 23. DOI: 10.1007/s43546-024-00521-2. URL: https://link.springer.com/article/10.1007/s43546-024-00521-2.

[5] Ziauddin Sami Sabouri and Behnam Mehrdel. "New Geopolitics of Artificial Intelligence and the Challenges of Global Governance". In: *Contemporary International Relations and Foreign Policy Journal* (2024). DOI: 10.30489/cifj.2024.431044.1094. URL: https://doi.org/10.30489/cifj.2024.431044.1094.

[6] Taylor Rodriguez Vance. *Geopolitical Implications of Artificial Intelligence in Cybersecurity: A Comprehensive Analysis*. Published on 04-August-2023. 2023. DOI: 10.5281/zenodo.8214594. URL: https://doi.org/10.5281/zenodo.8214594.

[7] Jiqiang Niu et al. "Global Research on Artificial Intelligence from 1990–2014: Spatially-Explicit Bibliometric Analysis". In: *ISPRS International Journal of Geo-Information* 5.5 (2016), p. 66. DOI: 10.3390/ijgi5050066. URL: https://doi.org/10.3390/ijgi5050066.

[8] Leo Schmallenbach, Till W. Bärnighausen, and Marc J. Lerchenmueller. "The global geography of artificial intelligence in life science research". In: *Nature Communications* 15 (2024), p. 7527. DOI: 10.1038/s41467-024-45187-1. URL: https://doi.org/10.1038/s41467-024-45187-1.

[9] Marcello M. Mariani et al. "Artificial intelligence in innovation research: A systematic review, conceptual framework, and future research directions". In: *Technovation* 122 (2023), p. 102623. DOI: 10.1016/j.technovation.2022.102623. URL: https://doi.org/10.1016/j.technovation.2022.102623.

[10] Ahmed H. Al-Marzouqi and Alya A. Arabi. "A Comparative Analysis of the Performance of Leading Countries in Conducting Artificial Intelligence Research". In: *Human Behavior and Emerging Technologies* (2024). DOI: 10.1155/2024/1689353. URL: https://doi.org/10.1155/2024/1689353.

[11] Jiajun Cao and Yuefen Wang. "International Cooperation Among Artificial Intelligence Research Teams Based on Regional Cooperation Models". In: *2020 ASIS&T Asia-Pacific Regional Conference (Virtual Conference), Wuhan, China, December 12-13, 2020*. Received: August 15, 2020; Accepted: September 15, 2020. 2020. DOI: 10.2478/dim-2020-0036. URL: https://doi.org/10.2478/dim-2020-0036.

[12] Vinayak, Adarsh Raghuvanshi, and Avinash Kshitij. "Signatures of capacity development through research collaborations in artificial intelligence and machine learning". In: *Journal of Informetrics* 17 (2023), p. 101358. DOI: 10.1016/j.joi.2023.101358. URL: https://www.sciencedirect.com/science/article/pii/S175115772300043X.

[13] Bedoor AlShebli et al. "Beijing's central role in global artificial intelligence research". In: *Scientific Reports* 12 (2022). Received: 02 March 2022; Accepted: 05 December 2022; Published: 12 December 2022, p. 21461. DOI: 10.1038/s41598-022-25714-0. URL: https://doi.org/10.1038/s41598-022-25714-0.

[14] Haotian Hu, Dongbo Wang, and Sanhong Deng. "Global Collaboration in Artificial Intelligence: Bibliometrics and Network Analysis from 1985 to 2019". In: *Journal of Data and Information Science* 5.4 (2020). Received: January 31, 2020; Revised: May 13, 2020; Accepted: June 11, 2020, pp. 86–115. DOI: 10.2478/jdis-2020-0027. URL: https://doi.org/10.2478/jdis-2020-0027.

[15] Gokhan Ozkaya and Ayse Demirhan. "Analysis of Countries in Terms of Artificial Intelligence Technologies: PROMETHEE and GAIA Method Approach". In: *Sustainability* 15.5 (2023), p. 4604. DOI: 10.3390/su15054604. URL: https://doi.org/10.3390/su15054604.

[16] *OpenAlex: A fully open catalog of the global research system*. https://openalex.org. Accessed: 2025-07-13. 2023.

[17] Stefano Baruffaldi et al. *Identifying and measuring developments in artificial intelligence: Making the impossible possible*. OECD Science, Technology and Industry Working Papers 2020/05. Paris: OECD Publishing, 2020. DOI: 10.1787/5f65ff7e-en. URL: https://doi.org/10.1787/5f65ff7e-en.

[18] Floriana Gargiulo et al. *A meso-scale cartography of the AI ecosystem*. 2022. arXiv: 2212.12263 [physics.soc-ph]. URL: https://arxiv.org/abs/2212.12263.

[19] Victor Bulatov, Vasiliy Alekseev, and Konstantin Vorontsov. "Determination of the Number of Topics Intrinsically: Is It Possible?" In: *Recent Trends in Analysis of Images, Social Networks and Texts. AIST 2023. Communications in Computer and Information Science*. Vol. 1905. Also available as arXiv preprint arXiv:2406.10402 [cs.CL]. Springer, Cham, 2024, pp. 1–15. DOI: 10.1007/978-3-031-67008-4_1. URL: https://doi.org/10.1007/978-3-031-67008-4_1.