Journal of Scientometric Research, 2026; 15(2):331-346.
https://www.jscires.org

Research Article

A Bibliometric Analysis of Explainable Artificial Intelligence (XAI): Trends, Themes, and Global Research Dynamics

Praveena Kebbekoppalu Shrinivasa[1], Mahadeva Swamy[1,*], Bhargavi Kalegowda[1], Pradeep Kumar Benakanahalli Puttappa[2]

[1]Department of ECE, Vidyavardhaka College of Engineering, Mysuru, Karnataka, INDIA.
[2]Department of Computer Science and Design, Atria Institute of Technology, Bengaluru, Karnataka, INDIA.

ABSTRACT

The active development of artificial intelligence resulted in the extensive introduction of complicated designs, most of them being black boxes. This has raised an urgent requirement of Explainable Artificial Intelligence (XAI) to enhance the explainability, trustworthiness, and interpretability of decision-making systems. The purpose of this research is to evaluate the current worldwide trends and patterns of research and development of XAI through a bibliometric approach. Major scientific databases with publications were searched, and this data was analyzed with the help of bibliometric tools and visualization techniques. The findings indicate that post 2019, there are intense growths in research production which point to the increasing interest in the topic. In terms of key words, artificial intelligence, machine learning and interpretability are prominent subjects of research. Analysis on a country level shows that the United States, the United Kingdom, and Germany contributed the most. The results also reveal the emergence of cross-functional fields of research, which integrate XAI with other fields that include psychology, medicine, and data science. According to the study, XAI research is slowly increasing in a transition towards the theoretical basis to the practical and application-based methods of research. These ideas allow getting a clear vision of the present day of research in explainable artificial intelligence and its further evolution.

Keywords: Bibliometric Analysis, Explainable AI, Interpretability, Machine Learning, Research Trends.

Correspondence:
Mahadeva Swamy
Department of ECE, Vidyavardhaka College of Engineering, Mysuru-570002, Karnataka, INDIA.
Email: mahadevaswamy@vvce.ac.in

Received: 13-04-2026;
Revised: 07-05-2026;
Accepted: 29-06-2026.

INTRODUCTION

The past few years have marked the swift development of Artificial Intelligence (AI), especially with its use in healthcare and finance, transportation and dozens of other fields (Jordan and Mitchell, 2015; LeCun et al., 2015). The accuracy and performance of modern AI systems, and in particular, deep learning systems, has been large. But the majority of these models are like black-box systems, where the inside decision-making process cannot be easily explained (Castelvecchi, 2016). Such a lack of transparency raises serious concerns, especially in high-stakes applications where the issues directly impact human lives (Rudin, 2019). Owing to this, explainability of AI systems is increasingly becoming required. Explainable Artificial Intelligence (XAI) is a model that seeks to explain AI (model prediction) to humans, enhancing transparency, trust and accountability (Adadi and Berrada, 2018; Guidotti et al., 2018). It is also significant to counter ethical issues of unfairness, prejudice, and trustworthiness (Samek et al., 2017). The need to have interpretable and reliable models has never been as essential as it is today as AI systems are incorporated more into real-world scenarios. Though several review papers have been released within the field of XAI, the majority are qualitative in sense targeting methods, techniques, or even certain applications (Arrieta et al., 2020; Doshi-Velez and Kim, 2017). The lack of thorough quantitative research that would consider the way the field develops over the years, what subjects become more significant, and what is the distribution of research partners worldwide remains. There have been many reviews in the research area on the methodologies for explainability, interpretability or domain specific applications, but most of the research conducted so far is qualitative and covers only methodological aspects instead of the overall development of the research area. Consequently, limited attention has been given to understanding the intellectual structure of the field, the patterns of scientific collaboration, the emergence of influential research themes, and the distribution of global scholarly contributions. The number of publications about XAI has grown by a large margin and traditional narrative reviews are less and less effective in summarizing the intersections between authors, institutions, countries, and research fields. Within this framework, bibliometric

DOI: 10.5530/jscires.20260179

analysis provides a scientifically sound and quantitative tool to represent the knowledge organization of a specific research field, define influential actors behind the publications, detect thematic developments and uncover emerging areas of research. In this regard, a thorough bibliometric evaluation of XAI is timely and much needed to present a comprehensive overview of the state of development of the XAI field, along with the dynamics and future path of research, within the time range 2016-2024 (Donthu et al., 2021; Zupic and Čater, 2015). Such analysis may provide an exact image of the current development of the field and flow of further research.

Besides looking at the perspectives of the present status, key issues and developments of the sector, this study will seek to answer the following research questions to present a holistic view of research of Explainable Artificial Intelligence:

- RQ1: What are the most impactful authors, journals, articles, countries, and organizations that contribute to the research in the field of Explainable Artificial Intelligence?
- RQ2: Which research themes are predominant in Explainable Artificial Intelligence? What has been the development of these themes and which are the commonest key words used?
- RQ3: What are the future opportunities and research directions in Explainable Artificial Intelligence based on thematic and trend analysis?

Therefore, the study paper aims to examine the development, composition, and changing of research on explainable artificial intelligence in a systematic manner through the application of bibliometric methods.

Related Works

Explainable Artificial Intelligence (XAI) is one of the most important research areas that seeks to tackle the opacity of complex machine learning systems. Artificial Intelligence (AI) systems are becoming more prevalent across critical sectors like healthcare, finance, and autonomous systems, making the need for transparency, interpretability, and trust more crucial than ever (Celepkolu *et al*., 2025). A literature review was conducted to combine the results of recent bibliometric and analytical studies on XAI and present an overview of trends, methodologies and gaps.

Evolution and growth of XAI research.

Throughout the studies, the growing field of XAI research is emphasized. Sharma *et al*. (2024) found that the number of publications has been rising over the years, especially from 2018 onwards, showing the increasing academic and industrial interest in explainability. Likewise, Celepkolu *et al*. (2025) also find that the number of publications has increased significantly since 2023, growing by around 62% by 2024, marking the growing significance of the field.

This expansion is also set against the backdrop of the development of the XAI literature between 1993 and 2024, shown by Russo (2026), which has undergone a shift from rudimentary understandings of XAI to a formal and multi-disciplinary body of study. Overall, the results indicate that XAI is a growing and unique field within AI.

Bibliometric Approaches and Research Mapping

A prevalent approach in studying the structure and dynamics of XAI research is bibliometric analysis. To understand the publication trend, citation pattern, and research clusters, Sharma *et al*. (2024) analyzed 4,781 publications from the Scopus database using the VOSviewer tool. Likewise, Celepkolu *et al*. (2025) analyse 13,961 records from the Web of Science, accounting for the temporal trends, geographical distribution and collaboration networks.

This is expanded by Russo (2026) who combines several bibliometric methods: citation analysis, co-citation networks and keyword co-occurrence mapping. This multi-dimensional approach allows to gain deeper insights in the intellectual, social, and conceptual structure of XAI research. The study highlights that the simultaneous application of these can offer a holistic picture of the evolution of research and the production of knowledge.

Key Research Themes and Clusters

The literature repeatedly mentions key clusters of themes in XAI. According to Celepkolu *et al*. (2025), XAI research can be divided into four main categories:

- Theoretical foundations.
- Methodological tools.
- System performance considerations.
- Application domains.

Similarly, Sharma *et al*. (2024) point out several new research directions, such as ethical considerations, legal aspects and explainability for humans. The study also highlights the need to consider biases, fairness, and accountability in AI systems.

Furthermore, Sharma *et al*. (2024) (Discover Applied Sciences paper) discuss various interdisciplinary uses of XAI, including in autonomous systems, cybersecurity, finance, and healthcare, showcasing its wider applicability. The results of these thematic analyses show that XAI research is far from a monolith, and is increasingly moving toward the social and ethical issues of society.

Patterns of collaboration and geographical distribution

Geographical and collaborative analyses show a focus of research in certain regions and research groups in XAI. The United States and China are identified by Celepkolu *et al.* (2025) as dominant contributors, where there are high AI institutional support and investments. Co-authorship network analyses also reveal that research collaborations tend to be grouped around prominent researchers and institutions, indicating influential research communities. This is echoed by an increase in the number of people who co-cited the same author in the same context, which he showed in Russo (2026) indicated the intellectual structure and the key figures defining the field.

Research on the Efficacy of XAI and Applications to Business and Industry

The reviewed studies show important methodological advances in the field of XAI. Russo (2026) reviews different interpretability techniques such as feature attribution methods, intrinsic and post-hoc explanations, and algorithmic unrolling techniques to improve interpretability without compromising the model's performance. Furthermore, there is a well-known balance between model complexity and interpretability highlighted by the literature. Although deep learning models achieve high performance, they are often not interpretable and thus require the creation of techniques to make them explainable (Celepkolu *et al.*, 2025). The challenges and research gaps are identified. Challenges and research gaps are identified. Despite immense advances, there are still some problems in the research of XAI. Sharma *et al.* (2024) highlight the limitations of their data sources, search strategies, and noting that potentially relevant studies may have been missed because of the use of keywords. Further, there is still a challenge to standardize evaluation measures for explainability. Ethical and legal issues are also a strong point of discussion. Fairness, accountability and adherence to regulations are highlighted in literature, including the right to explanation (Sharma *et al.*, 2024). Moreover, risks are mentioned of using unvalidated XAI techniques in sensitive areas such as healthcare, as mentioned by Russo (2026).

Future Research Directions

There is a need for further research on the standardization of the frameworks for evaluating, development of model-agnostic explainability techniques and better collaboration between different disciplines. Sharma *et al.* (2024) highlight the need for human-centric approach and ethical embedding within XAI systems. Furthermore, bibliometric analyses indicate that new areas of inquiry like fairness, transparency and trust are likely to remain prominent in research agendas.

METHODOLOGY

The paper is written in a rigorous bibliometric analysis framework, based on well-established scientometric methodologies and is structured into the following steps (Figure 1).

Database Selection

OpenAlex database was chosen as the main source of data used in the current study because there it provides a comprehensive coverage of scholarly publications in various fields. OpenAlex is open-access to large-scale bibliographic metadata, such as author details, keywords, citation formats, and affiliations, thus it can be appropriately used in large-scale bibliometric analysis (Priem et al., 2022). It is especially helpful to analyze emergent and quickly shifting areas of research like explainable artificial intelligence due to its extensive and constantly revising indexing.

Search Query Formulation

The search strategy was a keyword search and this was meant to make the various terminologies used with explainable artificial intelligence be captured. Search was carried out on titles, abstracts and keywords using Boolean operators and according to the standard practices of bibliometrics (Donthu et al., 2021). The last search query was:

TITLE-ABS-KEY (“Explainable AI” OR “XAI” OR interpretable machine learning” OR explainable AI OR explainability).

This practice also resulted in full coverage of the research area without much discrimination against useful literature.

Data retrieval and refinement

The preliminary search provided 3000 documents over the last five years between 2016 and 2024. Since OpenAlex aggregates bibliographic records from multiple scholarly sources, additional preprocessing and validation steps were carried out to ensure the reliability of the dataset. The retrieved records were first screened to identify and remove duplicate entries using DOI information, publication titles, and other bibliographic metadata. Samples with incomplete information or missing years of publication or inconsistent metadata were not included in further analysis. Author names, institutional affiliations and publication sources were reviewed and standardized as necessary for improved consistency throughout the dataset. Additionally, synonymous terms were combined, and common variations were harmonized for words associated with Explainable Artificial Intelligence (XAI) to do the normalization of the keywords. In addition, highly cited publications, authors, and sources were manually examined to verify the accuracy of citation and metadata information. These measures ensured that there were fewer potential problems arising from poor metadata quality (Inconsistencies, multiple authorships, different institutional affiliations, indexing errors etc.) and hence better quality and reliability of the bibliometric analysis (Aria and Cuccurullo, 2017). A final dataset of 1521 documents were derived after preprocessing (Figure 2).

Bibliometric Analysis

The filtered dataset was processed with a mixture of bibliometric methods which were installed via an analysis pipeline (written in Python). It has been analyzed using annual publication patterns to examine research expansions, key word co-occurrence analysis to determine prevalent themes in the research, and thematic mapping to categorize research themes in terms of centrality and development (Van Eck and Waltman, 2010). Analysis of abstract structure was done to find out clusters within the research field and analysis of collaboration network was done to find out relationship among countries and institutions. A citation analysis was also done to know the most influential authors, sources and organizations, which is a common practice in scientometric studies (Cobo et al., 2011).

Theme Identification and Interpretation

Clustering and thematic analyses were applied in identifying major research groups and new areas to know the intellectual structure of the field. New bonds were found between the publications and the research topics using bibliographic coupling and the keyword-based clustering techniques. The results of these reviews were categorized into highlighting key research areas, research directions, and future research prospects in explainable artificial intelligence as suggested in the previous bibliometric surveys (Farooq, 2024).

RESULTS AND DISCUSSION

This part includes the findings of the bibliometric study conducted on the research on explainable artificial intelligence. There were 3000 documents that were first of all drawn out of the database. Upon data cleaning, data filters and relevancy checks, 1521 documents were used to proceed with the final analysis. A summary is given in Table 1. The review targets to comprehend the expansion of publications, significant research themes, significant contributors and patterns of global collaboration. Different visualization models and statistical data help get a clear picture of the research field structure and development.

Growth of Publications and Citation Analysis

Two trends exhibiting a positive trend of growth in the period 2016 to 2024, the yearly scientific output of the research in the field of explainable artificial intelligence presents an upward trend. During the early years, the publications were extremely

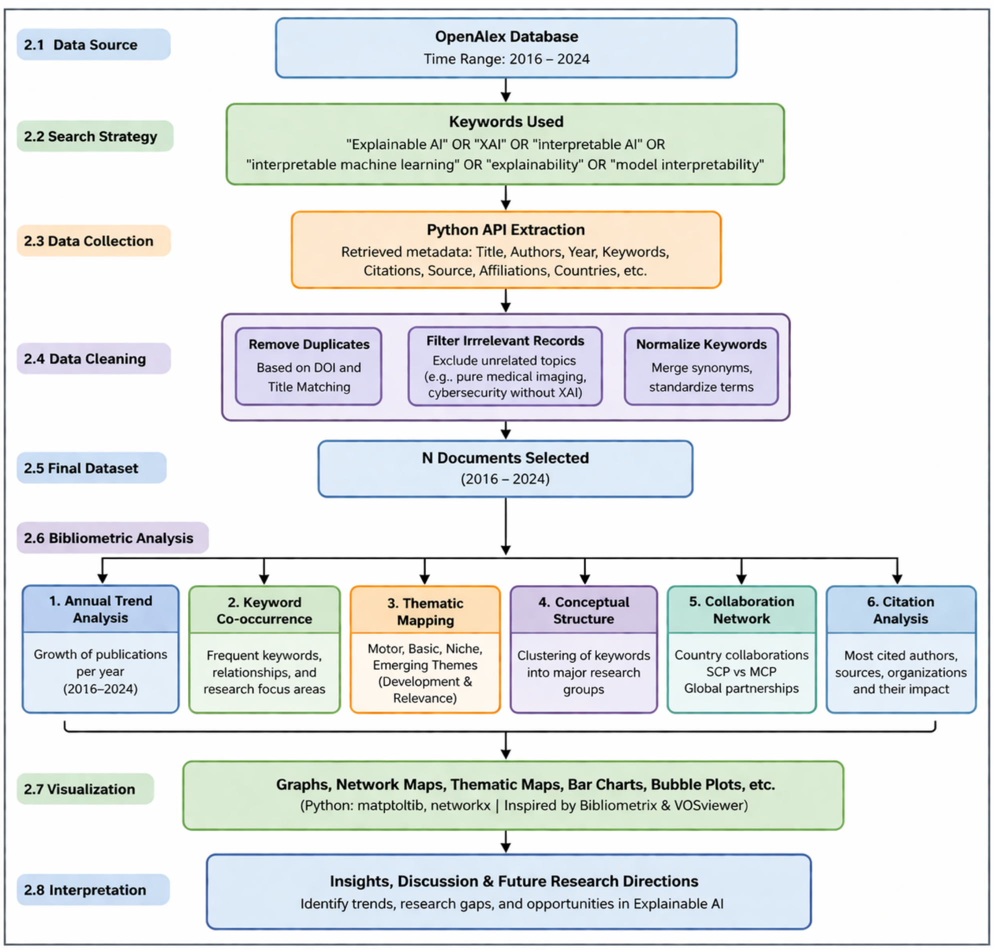


Figure 1: Overview of the bibliometric analysis framework adopted in this study.

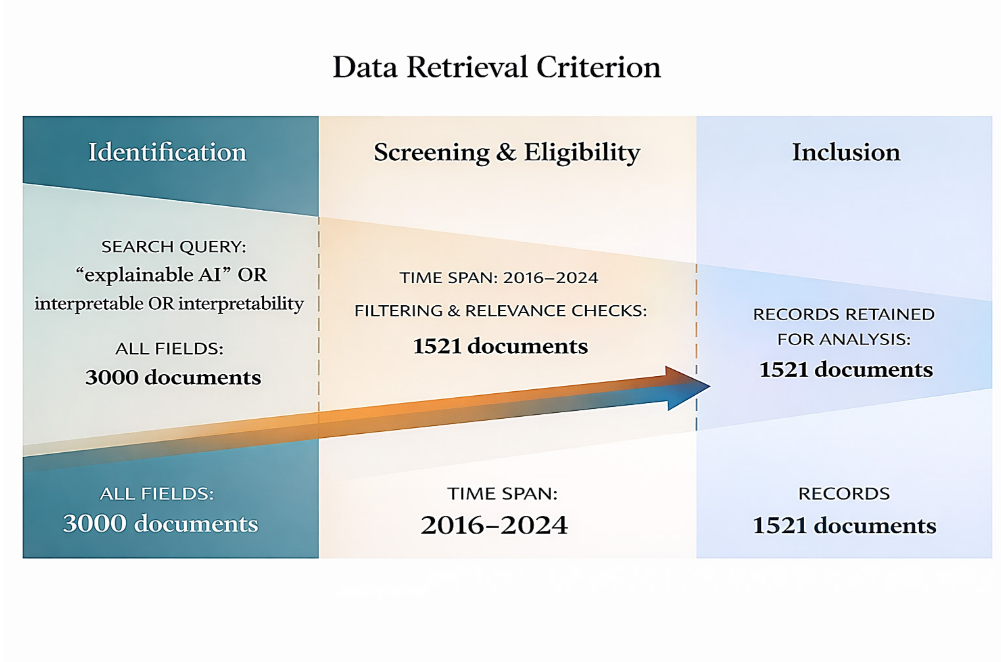


Figure 2: Data selection process.

Table 1: Summary of data.

Metric	Value
Total relevant papers	1521
Time span start	2016
Time span end	2025
Articles	1137
Reviews	119
Total citations	187648
Average citations per document	123.37
Average co-authors per document	4.51
No. of sources	589

small, only a small number of reports were published before 2018. Nevertheless, there is a slow upward trend since the year 2019 and suddenly after 2020 there is a steep increase. Its peak was in 2023 and 2024, which suggests a quick growth of activity in this field (Figure 3). This trend could be mostly ascribed to the rise in the use of artificial intelligence in essential applications, with model transparency and interpretability being the future needs. The increased interest in black-box models has motivated researchers to put more emphasis on explainability. It is important to note that the decreased number of the number of publications will be present in 2025, as not all the recent records were indexed, and it only signifies that there is no actual decrease in the level of research activity.

The most influential publications of the XAI domain based on citation count and total link strength are displayed in Table 2. First, the results of the survey show that the citation landscape is dominated by survey and review articles, reflecting a high need, within a fast-changing field, to consolidate and structure knowledge. Studies like the taxonomy proposed by Baredo Arrieta *et al.*, and the surveys by Adadi and Berrada, and Guidotti *et al.* have been encompassed as the basic references in the following research on XAI. Moreover, the existence of papers with high influence that speak about the importance of interpretable machine learning models indicates that progress is being made towards creating inherently transparent models, in contrast to just trying to explain black box models. The inclusion of model-agnostic approaches such as LIME and Anchors demonstrates the widespread adoption of explanation techniques that can be applied across diverse machine learning architectures. Further, the high number of citations for Grad-CAM++ or other neural network interpretation techniques highlights the growing significance of explainability in the application of deep learning. In general, the conclusions show three main research trends that push the evolution of XAI: conceptual framework and taxonomy, interpretable machine learning, and application of explanation methods for complex black-box models.

Table 3 shows the most influential authors in XAI according to publication volume, citation impact and collaboration strength. Findings indicate that the importance of scholarly contributions is clustered among a relatively small number of researchers that have significantly contributed to the evolution of explainability and interpretability methods. While Wojciech Samek shows the greatest number of publications and the largest collaboration network, citation impact is not necessarily linked to productivity- as the example of Javier Del Ser and Cynthia Rudin shows, they had similar number of citations despite having much fewer publications. Also, high overall link strength between leading authors suggests that collaborative research networks are crucial

Table 2: Most cited articles.

Sl. No.	Authors	Article Title	Citations	Total link strength
1.	Arrieta *et al.* (2020)	Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI	8402	520
2.	Rudin (2019)	Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead	8270	480
3.	Adadi and Berrada (2018)	Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)	5507	410
4.	Ribeiro *et al.* (2016)	"Why Should I Trust You?": Explaining the Predictions of Any Classifier	4905	450
5.	Guidotti *et al.* (2018)	A survey of methods for explaining black box models	4500	390
6.	Doshi-Velez and Kim (2017)	Towards A Rigorous Science of Interpretable Machine Learning	3119	360
7.	Chattopadhay *et al.* (2018)	Grad-CAM++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks	2904	300
8.	Montavon *et al.* (2018)	Methods for interpreting and understanding deep neural networks	2655	320
9.	Linardatos *et al.* (2020)	Explainable AI: A Review of Machine Learning Interpretability Methods	2609	280
10.	Ribeiro *et al.* (2018)	Anchors: High-Precision Model-Agnostic Explanations	2030	260

Table 3: Most cited authors.

Authors	Documents	Citations	Total link strength
Javier Del Ser	8	10571	84
Francisco Herrera	8	10568	80
Natalia Díaz-Rodríguez	11	10422	77
Wojciech Samek	24	10277	123
Cynthia Rudin	6	9943	14
Klaus-Robert Müller	15	9064	60
Adrien Bennetot	6	8596	45
Raja Chatila	5	8578	41
Siham Tabik	3	8537	31
Alberto Barbado	3	8456	23
Alejandro Barredo Arrieta	2	8437	22
Daniel Molina	2	8437	22
Richard Benjamins	2	8437	22
Salvador García	2	8437	22
Sergio Gil-López	2	8437	22
Sameer Singh	4	7367	9
Carlos Guestrin	3	7282	6
Grégoire Montavon	11	7195	44
Riccardo Guidotti	8	6867	62
Amina Adadi	2	5638	2

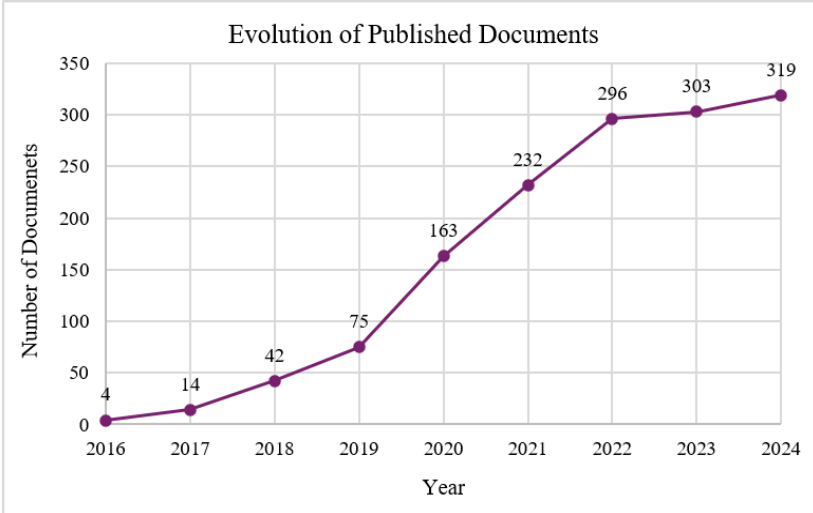


Figure 3: Evolution of Published documents.

to the innovation and dissemination of knowledge in the field. Several authors come from different research lines such as interpretable machine learning, explainability of neural networks, model-agnostic explainability, and XAI taxonomies, highlighting the multidisciplinary nature of XAI research. In addition, the result from recurring co-authorship patterns among several highly cited researchers indicates the presence of influential research clusters which have provided fundamental studies.

The following Table 4 shows the bibliometric impact of the leaders in XAI research based on the H-index, G-index, M-index, total citations, and publications. The findings show that Wojciech Samek and Klaus-Robert Müller exhibit the highest values of H-index and G-index, which means that they have the largest sustained scholarly impact. The higher the difference between the H-index and the G-index of an author, the more likely it is that there are a few outstanding publications that have had a big impact on the field. Moreover, higher M-index values are associated with the quick rate of citations and high visibility of the research in a relatively short period of time. Interestingly, several authors have appeared only a handful of times but have still had a strong impact on the number of citations they receive, emphasizing the importance of groundbreaking research and survey papers in XAI. The distribution of YSP values also shows the presence of early pioneers and newer contributors, reflecting the history of the field and its ongoing development. The results show that the development of XAI has been fueled by continued scholarly activity, through highly cited seminal papers, as well as by a group of emerging research leaders, in line with the growing maturity and consolidation of the field.

Explainable Artificial Intelligence (XAI) has a set of key sources that have significantly contributed to the development of the field, which are clustered in the publication landscape from Table 5. The journal titles like Information Fusion, IEEE Access and Nature Machine Intelligence underscore the robust influence of XAI in the larger field of AI and machine learning research. The high number of publications and network connections in the arXiv highlights the value of open-access dissemination platforms for fast dissemination of ideas and research advancements in this rapidly developing field. In addition, publication volume was not a significant factor in determining citation impact, as a handful of sources had high impact scores even though they published few XAI-related articles. Moreover, the impact scores of a few sources were very high even if they published few articles about XAI, which demonstrates that landmark contributions have more impact than the volume of publications. The interdisciplinary nature of the topics covered by the conference proceedings, review journals and the interdisciplinary venues suggests that XAI research is motivated by both the need for methodological innovation and the need for knowledge synthesis and that spreads its influence across application areas like Healthcare, Robotics, Cognitive Science and Signal Processing. All of these results show the development of a rich and varied publication landscape, enabling the ongoing development, sharing and cross-disciplinary uptake of research on EAI.

The review of impact analysis of publication sources (Table 6) indicates that development of Explainable Artificial Intelligence (XAI) has received a relatively limited number of influential publications sources and has attracted a large amount of scholarly interest. Investigating sources like arXiv, IEEE Access, or Lecture Notes in Computer Science, one finds high H-index, G-index and M-index values, reflecting the continued efforts of these sources to disseminate high-impact XAI research. Perhaps most importantly, arXiv shows the best bibliometric performance in terms of a number of metrics, and it has become increasingly relevant as a repository to support the rapid and efficient dissemination of knowledge in the AI community. Important differences with respect to the H-index and G-index values of some sources also indicate the existence of 'landmark publications' with a high impact on later research. Furthermore, there are very few papers published in XAI-related areas that have received such a high number of citations, with several papers in Nature Machine Intelligence and Information Fusion having very high citation impact. The inclusion of various types of venues (journals, conference proceedings, review venues, and interdisciplinary sources) highlights the multidisciplinary and diverse nature of XAI research.

The institutional distribution of XAI research from Table 7 shows that scientific influence is found among a relatively small number of universities, research centers and organizations that have made significant contributions in the development of the field from a technical perspective. Institutions with high citation counts, like the University of Washington, Duke University and the Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute have made significant contributions to the evolution of core XAI methods and applications. Interestingly, we find the strongest collaboration networks among organizations that also have high values for their total link strength, indicating that institutional influence is not just related to research output and impact of citations, but also to the active engagement in collaborative research activities. The fact that academic institutions and industrial research organizations were both represented reflects the interdisciplinary and application-oriented nature of XAI, with theoretical advancements being connected to real-world application problems. The geographical spread of the prominent organizations, ranging from North America, Europe, and Asia, also mirrors the global growth of the XAI field and the growing interest in creating AI systems that are transparent and trustworthy.

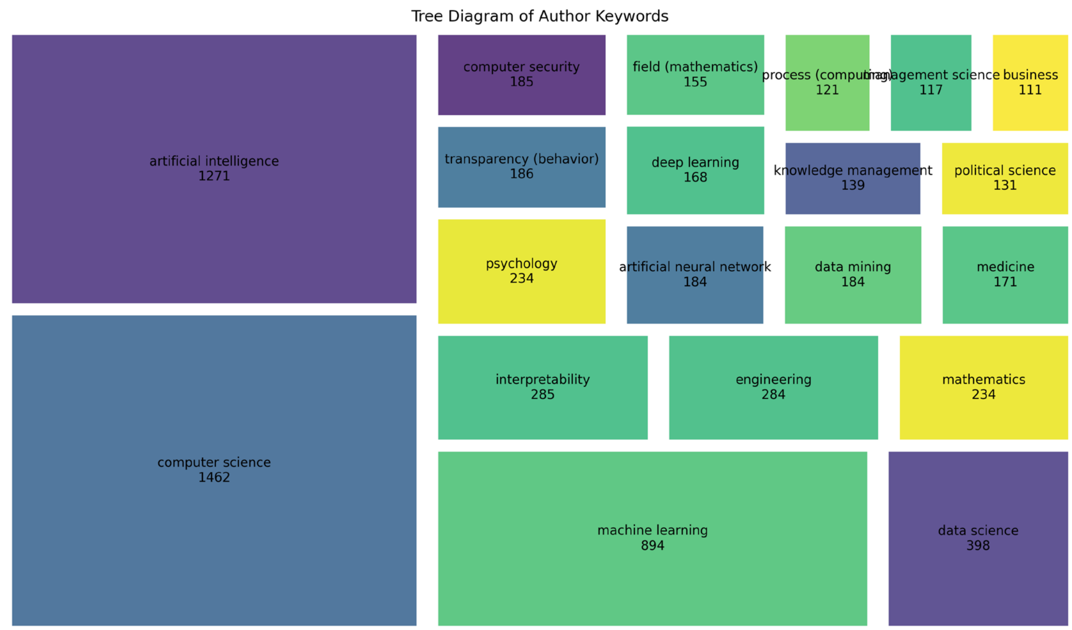
Keyword Analysis and Research Focus

The analysis of the author key words enables one to see the key themes of research in the world of explainable artificial intelligence. The most common and commonplace research activity as indicated in Figure 4 (a) are computer science and artificial intelligence. This is an easy indication that the discipline is well grounded with material computational disciplines. Also, the keywords like machine learning, data science, and interpretability are listed among others, and the peculiar emphasis is put on the development and explanation methods of the models. The existence of such terms as artificial neural networks, deep learning, and data mining is another manifestation of a major correlation between explainable AI and the new generation learning-oriented methods. Interestingly, a number of interdisciplinary keywords like psychology, medicine, political science, and management science are also discernible. This also indicates that explainable AI is becoming more frequently utilized in areas where people are vital in terms of their understanding, confidence, and judgment.

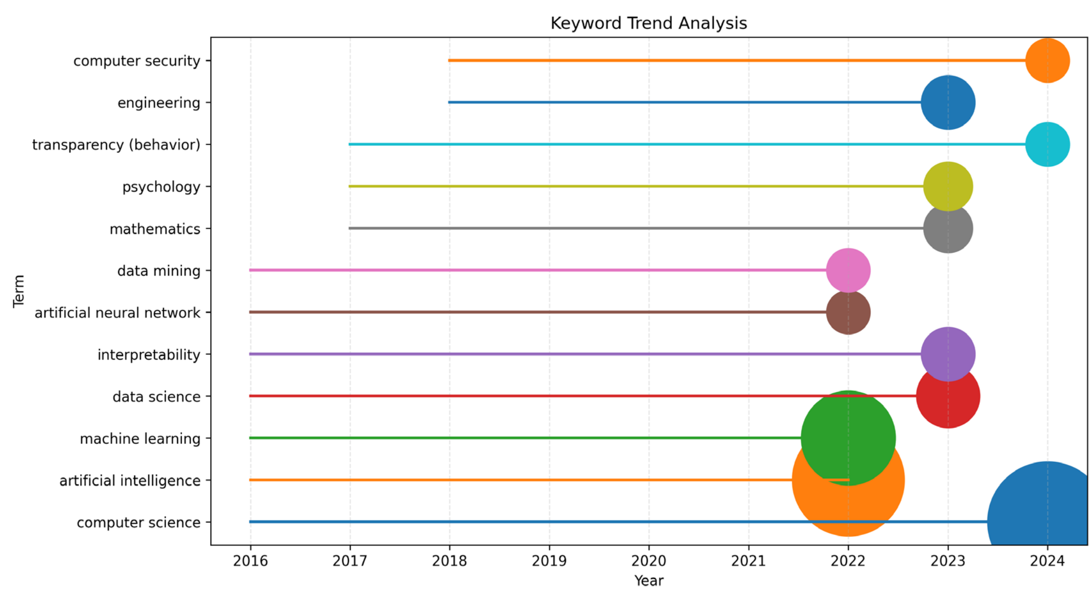
The trend analysis of keywords shows how the key research topics in explainable artificial intelligence are changing over time,

Table 4: Analysis of Author Impact.

Author	H-Index	G-Index	M-Index	TC	NP	YSP
Javier Del Ser	8	8	1	10571	8	2019
Francisco Herrera	8	8	1	10568	8	2019
Natalia Díaz-Rodríguez	11	11	1.375	10422	11	2019
Wojciech Samek	19	24	1.727	10277	24	2016
Cynthia Rudin	6	6	0.667	9943	6	2018
Klaus-Robert Müller	14	15	1.273	9064	15	2016
Adrien Bennetot	6	6	0.75	8596	6	2019
Raja Chatila	5	5	0.625	8578	5	2019
Siham Tabik	3	3	0.375	8537	3	2019
Alberto Barbado	3	3	0.375	8456	3	2019
Alejandro Barredo Arrieta	2	2	0.25	8437	2	2019
Salvador García	2	2	0.25	8437	2	2019
Sergio Gil-López	2	2	0.25	8437	2	2019
Daniel Molina	2	2	0.25	8437	2	2019
Richard Benjamins	2	2	0.25	8437	2	2019
Sameer Singh	4	4	0.364	7367	4	2016
Carlos Guestrin	3	3	0.273	7282	3	2016
Grégoire Montavon	10	11	0.909	7195	11	2016
Riccardo Guidotti	8	8	1	6867	8	2019
Amina Adadi	2	2	0.222	5638	2	2018



(a)



(b)

Figure 4: Keyword trend analysis (a) Treemap visualization of Author Keywords (b) trend over the years (Source: Vosviewer).

as depicted in Figure 4 (b). At the dawn of times, the study was mainly focused on the traditional areas like computer science and artificial intelligence which became the keystones to the discipline. Since approximately 2018, the keywords have also shifted towards model development and data-driven solutions, evident in keywords like machine learning, artificial neural networks, and data mining. This is indicative of the increased use of learning-based approaches in artificial intelligence studies. In more recent times, especially since 2020, the terms of interpretability, data science and transparency have become very prominent. The given trend underlines the growing importance of studying and describing complex models in combination with their use in the environment where trust and responsibility play a major role.

Thematic Structure of the Research Field

The thematic map allows us to gain a systematic perspective of research space in explainable artificial intelligence by centrality and development. Figure 5 shows that the research themes fall in four quadrants i.e. the motor themes, the basic themes, the niche themes and emerging or disappearing themes. The motor themes are in the upper-right quadrant; these areas are

Table 5: Most cited sources.

Source	Documents	Citations	Total link strength
Information Fusion	16	13454	520
arXiv (Cornell University)	99	11155	610
IEEE Access	44	9319	540
Nature Machine Intelligence	2	8355	470
ISTI Open Portal	1	4500	180
Lecture notes in computer science	38	4497	500
Proceedings of the AAAI Conference on Artificial Intelligence	10	3664	460
Wiley Interdisciplinary Reviews Data Mining and Knowledge Discovery	6	3016	320
Digital Signal Processing	1	2655	250
Entropy	1	2609	240
IEEE Transactions on Neural Networks and Learning Systems	5	2283	390
ACM Computing Surveys	12	2268	410
Proceedings of the National Academy of Sciences	1	1993	260
Artificial Intelligence Review	7	1988	350
Pattern Recognition	4	1950	330
BMC Medical Informatics and Decision Making	5	1854	300
Science Robotics	1	1817	210
Journal of Artificial Intelligence Research	4	1735	360
Cognitive Computation	5	1697	280
Artificial Intelligence	8	1515	370

well-documented and of great significance in research. These incorporate issues pertaining to interpretability, transparency, as well as fundamental artificial intelligence uses. The location of the themes shows that not only are they the core of the field, but they are in active development, which is the conduit of the current research. The fundamental ones, located at the lower-right quadrant, can primarily be described as the areas of basic understanding like computer science, and general phenomena of artificial intelligence. These themes are crucial to the discipline, but which are not as well-articulated in terms of specialized study as are the motor themes. The theme in the upper-left quadrant is the niche ones, which cover things like machine learning, mathematics, and artificial neural networks. These fields are well advanced but are rather secluded which means that they have great technical in-depths but little integration with the rest of interdisciplinary applications. The rising or fading ones can be found on the bottom-left quadrant, and they comprise data science, psychology, and data mining. These themes are either new strands of research being pursued, or those that are slowly falling out of favor. Here they imply a movement towards interdisciplinarity exploration, in particular the human-centered facets of explainable AI.

The network visualization and conceptual structure map can give a deeper insight into the rational arrangement of the explainable artificial intelligence research as depicted in Figures 6 and 7. The conceptual structure map demonstrates that there are several different clusters, which means that the sphere of research is segmented into several thematic groups. There is one cluster (primarily linked to core ideas of artificial intelligence and machine learning) which is the technical foundation of the field. Application-oriented research is represented in one more cluster, in which explainable AI methods are used in the practical issues. Another third cluster that is connected to interpretability and human-centered considerations deals with helping people understand AI models and trust them, as well as being transparent. The fact that there is a distance between clusters is an indication that these directions of research are growing side by side. Nevertheless, there is a growing approach between the application-based research and technical-driven research due to the relative proximity of some of these clusters. The network visualization also helps to confirm this and emphasize the relations between various keywords and research topics. Heavy edges portray common ideas that might be used together suggesting high thematic links among networks. Central nodes are associated with central areas of artificial intelligence, machine learning, and interpretability in bridging with different subfield.

The nation work network of collaboration demonstrates that explainable AI research has an extremely extensive network at the international scale. The United States becomes the hub center but closely interconnected to a few countries such as

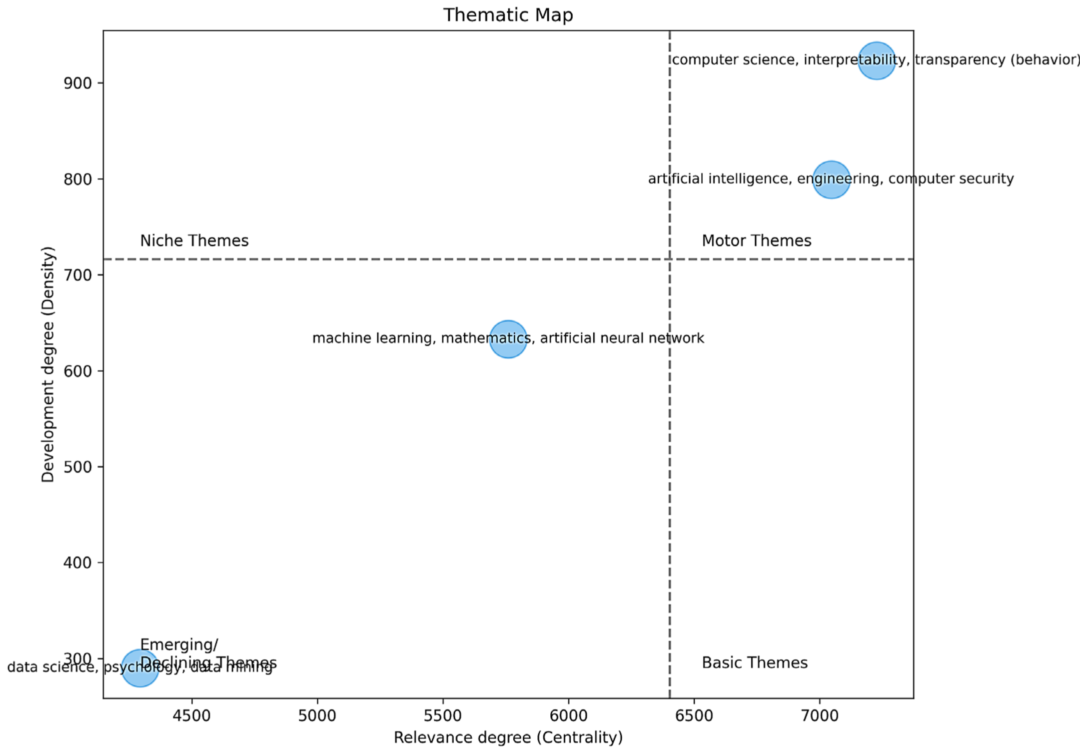


Figure 5: Thematic map (Source: Biblioshiny).

Table 6: Analysis of most impact sources.

Source	H-Index	G-Index	M-Index	TC	NP	YSP
Information Fusion	16	16	2	13454	16	2019
arXiv (Cornell University)	39	99	3.545	11155	99	2016
IEEE Access	31	44	3.444	9319	44	2018
Nature Machine Intelligence	2	2	0.25	8355	2	2019
ISTI Open Portal	1	1	0.125	4500	1	2019
Lecture notes in computer science	24	38	2.667	4497	38	2018
Proceedings of the AAAI Conference on Artificial Intelligence	9	10	1	3664	10	2018
Wiley Interdisciplinary Reviews Data Mining and Knowledge Discovery	6	6	0.75	3016	6	2019
Digital Signal Processing	1	1	0.1	2655	1	2017
Entropy	1	1	0.143	2609	1	2020
IEEE Transactions on Neural Networks and Learning Systems	5	5	0.714	2283	5	2020
ACM Computing Surveys	12	12	2.4	2268	12	2022
Proceedings of the National Academy of Sciences	1	1	0.125	1993	1	2019
Artificial Intelligence Review	7	7	1	1988	7	2020
Pattern Recognition	4	4	0.364	1950	4	2016
BMC Medical Informatics and Decision Making	5	5	0.714	1854	5	2020
Science Robotics	1	1	0.125	1817	1	2019
Journal of Artificial Intelligence Research	4	4	0.667	1735	4	2021
Cognitive Computation	5	5	1.25	1697	5	2023
Artificial Intelligence	8	8	1.143	1515	8	2020

*TC-Total Citations *NP- Number of Publications *YSP- Year of Starting Publication.

Table 7: Most influential organizations.

Organization	Documents	Citations	Total link strength
University of Washington	19	10252	225
Duke University	6	9299	65
Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute	19	9156	530
Sorbonne Université	10	8904	290
École Nationale Supérieure de Techniques Avancées	3	8533	75
Institut Systèmes Intelligents et de Robotique	4	8527	95
Tecnalia	1	8402	25
Telefonica Research and Development	1	8402	25
École Nationale Supérieure de Techniques Avancées Paris	1	8402	25
University of Pisa	19	7646	360
Technische Universität Berlin	19	7455	465
Korea University	18	7168	395
Sidi Mohamed Ben Abdellah University	2	5638	20
Istituto di Scienza e Tecnologie dell'Informazione "Alessandro Faedo"	6	4642	70
Max Planck Institute for Informatics	8	4067	205
Medical University of Graz	18	4026	425
ETH Zurich	16	3066	285
Defense Advanced Research Projects Agency	3	3024	40
Indian Institute of Technology Hyderabad	3	2940	55
University of Patras	4	2923	40

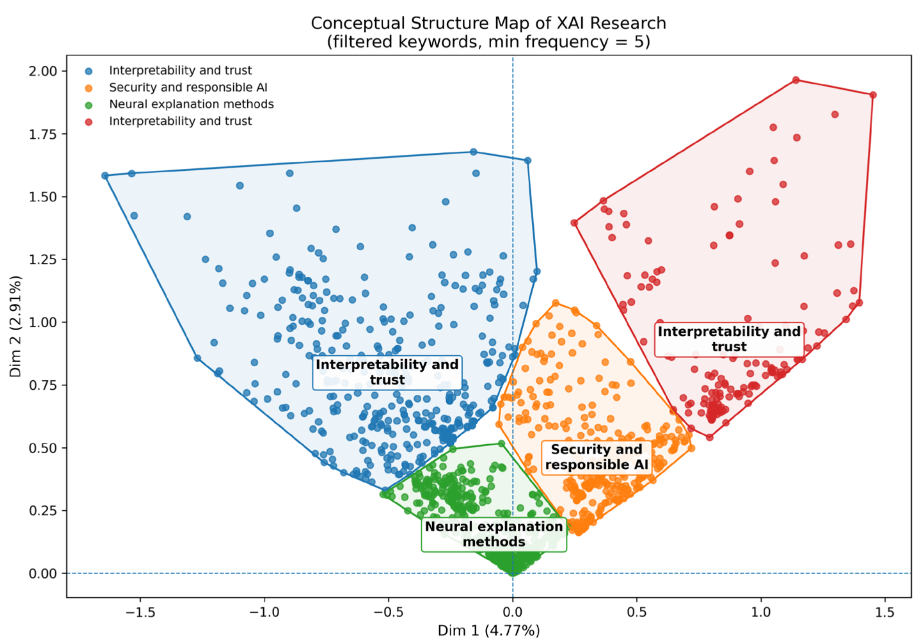


Figure 6: Conceptual Structure Map (Source: Biblioshiny).

the United Kingdom, Germany and China. Interconnections are also strong in European countries that create a great net of collaboration. Conversely, the developing nations like India are also actively contributing but are not as at the centre point of co-operation of the world fabric. The multi-country analysis on production further suggests that the level of Multi Country Publications (MCP) is much higher in case of leading countries, which translates the high value of international collaboration to enhance research in this area.

Table 8 shows the scientific production analysis in terms of the leadership and collaboration patterns in various countries. The findings show that the United States is the largest producer of research with their contribution of 373 documents, and there is an equal split between Single-Country Publications (SCP) and Multiple-Country Publications (MCP). It is indicative of good independent research skills and proactive international cooperation. The United Kingdom, Germany, and China also show considerable contribution in terms of research and their MCP values are relatively high, which means that they have significant participation in collaborative research. Specifically, the United Kingdom has demonstrated a stronger rate of international collaborations and this is indicated by the MCP ratio of 0.659. India seems to be one of the leading contributors that have a balanced distribution of SCP, and MCP, which might demonstrate a combination of a strong domestic research base and a growing international cooperation. In the same vein, other nations such as Italy, Australia and South Korea also have a significant contribution with a high MCP ratio (0.753) which reflects on their strong reliance on international alliances. Interestingly, other countries like Saudi Arabia and Malaysia have very high MCP ratios (more than 0.85) which indicates that their research output is more fuelled by international cooperation and not individual input. Ireland, in contrast, is lower in MCP ratio (0.367), suggesting a higher focus on conducting research at a single country.

The three-field diagram (Figure 8) emphasizes the topic of interaction between leading countries, prevailing keywords and significant places of publication. We see that the United States, United Kingdom and Germany are the leading contributors in various research themes especially in computer science, artificial intelligence and machine learning. These topics also positively overlap with journals of high impact like IEEE access, Information fusion and Lecture Notes in Computer Science, which shows that essential research carried out by the researcher is highly characterized by well-established journals. New competitors like India and China are increasingly engaged in a variety of major research areas especially in the field of data science and engineering with an indication that they will be increasingly involved in the field. The patterns of linkages clearly show that the major nations are not confined to certain subjects as they are playing a major part in setting contribution in various fields.

These findings are also supported by the collaborative production leadership distribution (Figure 9) which shows the equilibrium between Single-Country Publications (SCP) and Multiple-Country Publications (MCP). The United States is exhibiting leadership that has high SCP values meaning it has strong capability of independent research. Conversely, other nations including the United Kingdom, Australia and France have greater MCP proportions indicative of good international network of collaboration. The presence of countries such as Saudi Arabia and Malaysia, with leading MCP contributions, means that they rely heavily on global research affiliations. At the same time, India and Germany have a balanced prevalence of SCP and

Table 8: Analysis according to scientific leadership and the type of collaboration.

Rank	Country	Documents	SCP	MCP	Freq.	MCP_Ratio
1	United States	373	194	179	0.245	0.48
2	United Kingdom	167	57	110	0.11	0.659
3	Germany	151	70	81	0.099	0.536
4	India	148	76	72	0.097	0.486
5	China	137	63	74	0.09	0.54
6	Italy	87	34	53	0.057	0.609
7	Australia	77	19	58	0.051	0.753
8	South Korea	77	28	49	0.051	0.636
9	Canada	66	27	39	0.043	0.591
10	Spain	54	19	35	0.036	0.648
11	France	52	13	39	0.034	0.75
12	Saudi Arabia	51	6	45	0.034	0.882
13	Bangladesh	50	20	30	0.033	0.6
14	Netherlands	48	15	33	0.032	0.688
15	Austria	41	9	32	0.027	0.78
16	Switzerland	41	12	29	0.027	0.707
17	Japan	31	13	18	0.02	0.581
18	Ireland	30	19	11	0.02	0.367
19	Malaysia	28	3	25	0.018	0.893
20	Norway	26	8	18	0.017	0.692

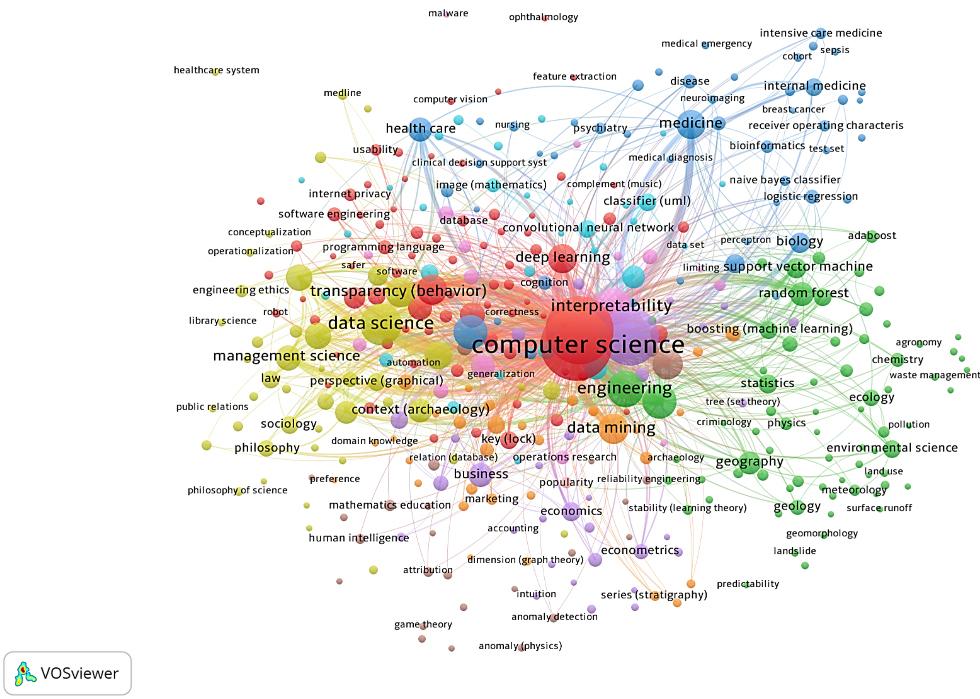


Figure 7: Network visualization (Source: Vosviewer).

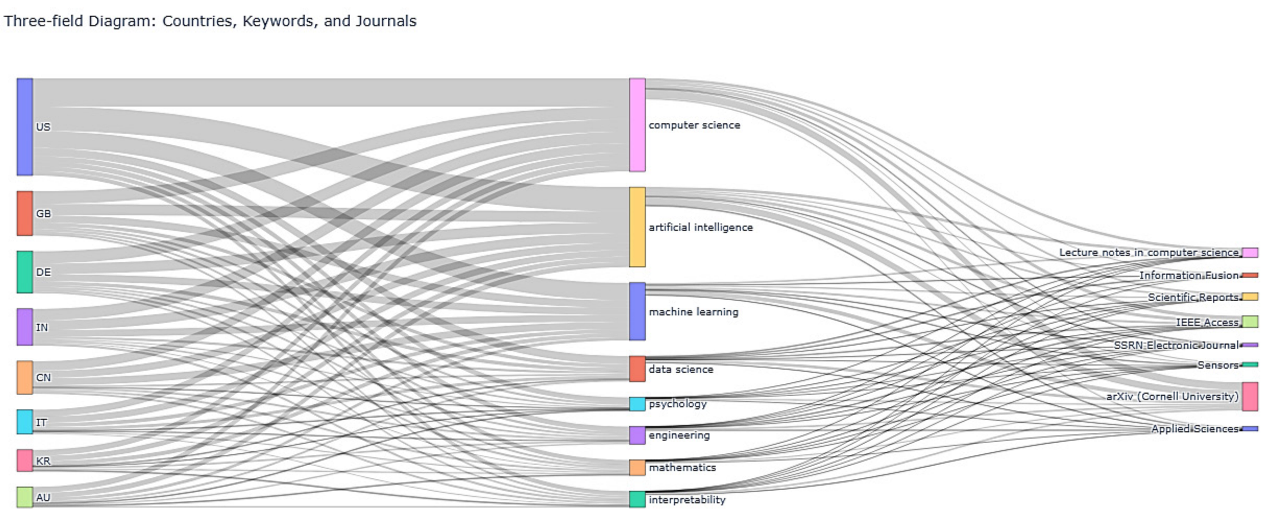


Figure 8: Three-field diagram Countries, Keywords, and Journals (Source: Biblioshiny).

MCP, which implies a balance between domestic and foreign cooperation.

LIMITATIONS

This paper offers bibliometric reviews of the research space of explainable artificial intelligence based on data gathered within the scope of significant academic databases. Although the analysis is systematic, data-driven, some limitations must be mentioned. First, these data are restricted to those publications indexed in certain databases, so all of the studies that could have been used were not considered. This can have an insignificant impact on the general statistics of coverage of the analysis. Second, the researchers limit the research to articles and reviews that were published in English. This has led to the possibility of involving vital contributions to be published in other languages. Third, the extraction of the data has been conducted at a certain time, and the research area is constantly changing. Even though the overall trends are likely to stay the same, some of the recent trends might not be fully reflected. Fourth, as the main

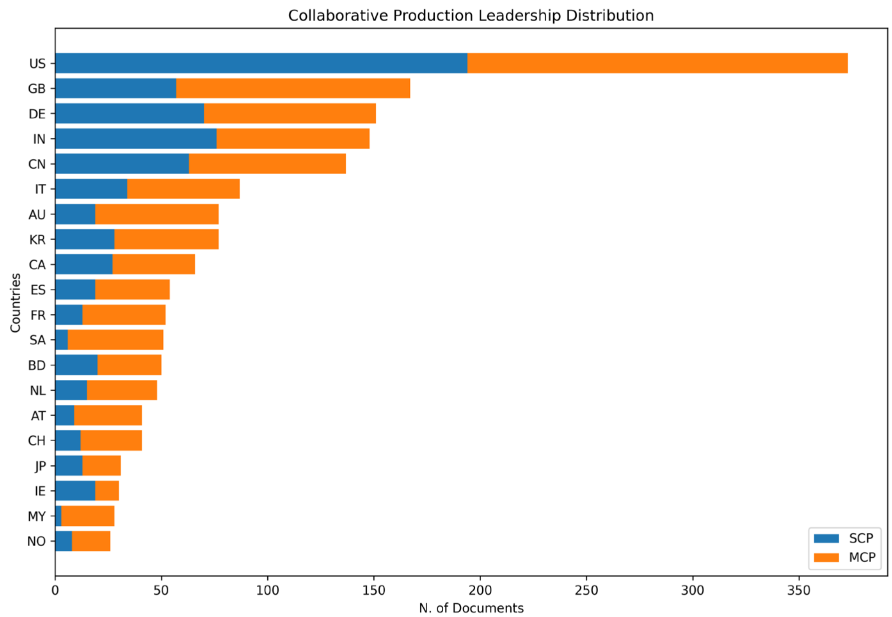


Figure 9: Collaborative production leadership distribution (Source: Biblioshiny). *SCP - Single Country Publications, * MCP - Multiple Country Publications.

measure of analysis; bibliometric indicators including keywords, citations and metadata are the foundation of the analysis. Such a method can be quite insensitive to the fine details of context and methodological depth of single research. Despite these weaknesses, the research provides an excellent overview of the organization and development of explainable AI studies. Future research that integrates more databases and enlarged data sources will allow getting a deeper insight into the domain.

CONCLUSION

This paper, based on bibliometric analysis of works in the area of explainable artificial intelligence, provides a clear account of the way the field has evolved throughout the years. The findings indicate that the field primarily consists of building upon computer science and artificial intelligence but shifts towards issues of interpretability, transparency, and trust. The keyword and trend analysis reveals that previous work concentrated more on how to construct a model, yet the current work is centering more on how to understand and explain such models. According to the thematic and conceptual analysis, the field is not restricted to some technical fields per se, but gradually establishing contact with other fields like psychology, medicine, and management. The collaboration findings indicate that the presence of both isolated research and global partnerships in some countries such as the United States and Germany is robust whereas there are other countries that are more reliant on collaborations. The three-field analysis indicates as well that all the main countries produce in various subjects and presenting in a famous journal. The research indicates that the discipline is shifting to do applications, wherein knowledge and confidence in models are needed.

Future bibliometric studies on XAI can be expanded by incorporating multiple databases, such as Web of Science, Scopus, and Dimensions, to provide broader coverage and improve result reliability. Comparisons from these databases can also identify differences in publication trends and the forms of citations. Future research may also focus on the evolution of topics, the Science Citation Network Search, and Science Frontiers identification through sophisticated science-mapping methods. XAI is a field that is rapidly evolving, making it crucial to conduct periodic bibliometric updates to identify promising trends, top-tier authors and new application fields covering topics like generative AI, large language models, and responsible AI.

ABBREVIATIONS

AI: Artificial Intelligence; **XAI:** Explainable Artificial Intelligence; **SCP:** Single-Country Publications; **MCP:** Multiple-Country Publications; **TC:** Total Citations; **NP:** Number of Publications; **YSP:** Year of Starting Publication; **DOI:** Digital Object Identifier; **Grad-CAM++:** Generalized Gradient-Based Visual Explanations; **RQ:** Research Question.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

REFERENCES

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access*, *6*, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

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

Barredo Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Castelvecchi, D. (2016). Can we open the black box of AI? *Nature*, *538*(7623), 20–23. https://doi.org/10.1038/538020a

Celepkolu, K. C., Karaoglan, M., & Turker, T. (2025). A bibliometric analysis of explainable artificial intelligence research. In *Proceedings of the international conference on recent academic studies*.

Chattopadhay, A., Sarkar, A., Howlader, P., & Balasubramanian, V. N. (2018). Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks IEEE winter conference on applications of computer vision (WACV). https://doi.org/10.1109/WACV.2018.00097

Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). Science mapping software tools: Review, analysis, and cooperative study among tools. *Journal of the American Society for Information Science and Technology*, *62*(7), 1382–1402. https://doi.org/10.1002/asi.21525

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

Doshi-Velez, Finale, & Kim, B. *"Towards a rigorous science of interpretable machine learning." arXiv preprint arXiv:1702.08608*. (2017).

Farooq, R. (2024). A review of knowledge management research in the past three decades: A bibliometric analysis. *VINE Journal of Information and Knowledge Management Systems*, *54*(2), 339–378. https://doi.org/10.1108/VJIKMS-08-2021-0169

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM Computing Surveys*, *51*(5), 1–42. https://doi.org/10.1145/3236009

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

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning." nature 521.7553. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2020). Explainable ai: A review of machine learning interpretability methods. *Entropy*, *23*(1), Article 18. https://doi.org/10.3390/e23010018

Lundberg, S. M., & Su-In Lee. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*.

Molnar, C. Interpretable machine learning. *Lulu. com*. (2020).

Montavon, G., Samek, W., & Müller, K.-R. (2018). Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*, *73*, 1–15. https://doi.org/10.1016/j.dsp.2017.10.011

Priem, J., Piwowar, H., & Orr, R. *"OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts." arXiv preprint arXiv:2205.01833*. (2022).

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 97–101). Association for Computational Linguistics. https://doi.org/10.18653/v1/N16-3020

Ribeiro, M. T., Singh, S., & Guestrin, C. (2018). Anchors: High-precision model-agnostic explanations. *Proceedings of the AAAI Conference on Artificial Intelligence*, *32*(1). https://doi.org/10.1609/aaai.v32i1.11491

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

Russo, M., & Vistocco, D. (2026). Explainable artificial intelligence through the lens of bibliometric citation analysis. *Applied Stochastic Models in Business and Industry*, *42*(3). https://doi.org/10.1002/asmb.70091

Samek, W., Wiegand, T., & Müller, K.-R. *"Explainable artificial intelligence: understanding, visualizing and interpreting deep learning models." arXiv preprint arXiv:1708.08296*. (2017).

Sharma, C., Sharma, S., Sharma, K., Sethi, G. K., & Chen, H.-Y. (2024). Exploring explainable AI: A bibliometric analysis. *Discover Applied Sciences*, *6*(11), 615. https://doi.org/10.1007/s42452-024-06324-z

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

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

Cite this article: Shrinivasa PK, Mahadeva Swamy M, Kalegowda B, Puttappa PKB. A Bibliometric Analysis of Explainable Artificial Intelligence (XAI): Trends, Themes, and Global Research Dynamics. J Scientometric Res. 2026;15(2):331-46.