Babylonian Journal of Artificial Intelligence
Vol. **(2025)**, 2025, **pp**. 37–54
DOI: https://doi.org/10.58496/BJAI/2025/004 ;ISSN: 2958-6453
https://mesopotamian.press/journals/index.php/BJAI

Research Article

Analyzing Trustworthy in AI: A Comprehensive Bibliometric Review of Artificial Intelligence Research

Klodian Dhoska [1,2,*], Elena Bebi [3]

[1] *Mechanical Department, Polytechnic University of Tirana, Albania.*
[2] *Association of Talent under Liberty in Technology (TULTECH), Tallinn, Estonia.*
[3] *Faculty of Mechanical Engineering, Polytechnic University of Tirana, Albania.*

ARTICLE INFO

Article History
Received 10 Jan 2025
Revised: 07 Feb 2025
Accepted 05 Mar 2025
Published 31 Mar 2025

Keywords
Trustworthy AI
bibliometric analysis
Artificial intelligence

ABSTRACT

This bibliometric analysis delves into the landscape of Trustworthy Artificial Intelligence (AI) research, revealing intriguing patterns and key insights. Lotka's Law unveils a skewed distribution in author productivity, underscoring the concentration of scholarly output among a select few. Bradford's Law identifies core journals significantly contributing to the field's scientific productivity. Author affiliations shed light on influential institutions, with Tsinghua University and Beijing University of Posts and Telecommunications emerging as major contributors. Examining corresponding authors' countries emphasizes China's dominance in both citations and hosting corresponding authors. Authors with an h-index of 9 and above showcase local impact, with researchers like LI J and ZHANG J standing out. The Collaboration Network visualizes the interconnectedness of researchers, revealing collaborative clusters. Analyzing countries' scientific production underscores China's leadership, with a substantial global contribution. Noteworthy documents and their impact, such as LIU R's work garnering high total citations, illustrate the significance of specific publications. These findings underscore the global and multifaceted nature of Trustworthy AI research, providing valuable insights for future investigations, policy considerations, and international collaborations in this rapidly evolving field.

1. INTRODUCTION

In recent years, the exponential growth of Artificial Intelligence (AI)[1] has given rise to a myriad of advancements across diverse domains, underscoring the importance of Trustworthy AI[2]. Trustworthy AI, characterized by attributes such as reliability, transparency, and ethical considerations, is pivotal for the responsible development and deployment of AI technologies. This study employs bibliometric analysis as a lens to comprehensively explore the landscape of Trustworthy AI research[3].

Bibliometrics, as a quantitative analysis of scholarly publications, provides a robust methodology for systematically evaluating the scholarly output, key contributors, collaborative networks, and global impact of research in the Trustworthy AI domain[4]. In this exploration, several overarching questions emerge. How is author productivity distributed in Trustworthy AI research, and what patterns emerge in the scholarly output? What are the core sources that significantly contribute to the dissemination of Trustworthy AI knowledge, and how does this align with bibliometric principles? How do local impacts of authors, affiliations, and corresponding author's countries influence the trajectory of Trustworthy AI research on a broader scale? What regional and global dynamics characterize the scientific production and impact of countries in this field? How does collaboration manifest in the Trustworthy AI research network, and what insights does it offer into the interconnectedness of researchers?

This study unfolds across various sections, each contributing a unique perspective to the broader understanding of Trustworthy AI research. We begin by examining general patterns in author productivity, exploring the overall impact on the scholarly output in this field. The concentration of impactful sources is investigated to reveal core journals that significantly contribute to Trustworthy AI knowledge dissemination. An exploration of authors' local impact, affiliations, and corresponding author's countries provides insights into the contextual factors influencing research outcomes. The examination of global production and citation impact of countries sheds light on regional variations and the influence of geopolitical factors on scientific contributions. Finally, the collaboration network analysis visualizes the intricate web of relationships among researchers, offering insights into the collaborative dynamics shaping Trustworthy AI research.

*Corresponding author. Email: kdhoska@fim.edu.al

Through these lenses, we aim to paint a comprehensive picture of Trustworthy AI research, providing valuable insights for scholars, practitioners, and policymakers alike. This journey into the bibliometric landscape of Trustworthy AI offers not only a snapshot of the current state but also lays the groundwork for future explorations, pushing the boundaries of knowledge in this critical domain.

2. METHODOLOGY

2.1 Database Selection:

We conducted a comprehensive bibliometric analysis focusing on the theme of trustworthiness in artificial intelligence (AI) research. Scopus was chosen as the primary database for its extensive coverage of scholarly publications in the fields of computer science and decision science.

2.2 Search Query:

The search query employed for database retrieval was designed to capture relevant articles on trustworthiness and reliability in AI. The keywords "Trustworthiness," "Reliability," and "Trustworthy" were combined with "AI" and "Artificial Intelligence." The search was targeted specifically at the article title, keywords, and abstract to ensure the relevance of the retrieved documents.

2.3 Inclusion Criteria and Filters:

To narrow down the search results, we applied specific filters:
Year Range: The publication years were restricted to the period from 2018 to 2024.
Subject Areas: The search was confined to articles falling within the subject areas of computer science and decision science.

2.4 Resulting Dataset:

The search strategy yielded a total of 4985 documents meeting the inclusion criteria. These documents constitute the basis for our bibliometric analysis of trustworthiness in AI research.

2.5 Data Extraction and Analysis:

RStudio, utilizing the R programming language, was employed for data extraction and analysis. The biblioshiny package in R facilitated the extraction of figures and tables, allowing for a comprehensive overview of the bibliometric data.

2.6 Completeness of Bibliographic Metadata:

To assess the completeness of bibliographic metadata, various metadata elements were evaluated. The evaluation revealed that most metadata, including author information, document type, language, publication year, title, and total citations, was complete. However, some elements, such as journal source (SO) and abstract (AB), exhibited minor gaps, while others like DOI (DI), keywords (DE), and keywords plus (ID) showed a higher degree of incompleteness. Despite these challenges, we mitigated issues arising from missing metadata and proceeded with the analysis based on the available information.

2.7 Metadata Quality Assessment:

A qualitative assessment was conducted to categorize the completeness of metadata into different levels, ranging from "Excellent" to "Poor" and "Completely Missing." This assessment guided our understanding of potential limitations in the dataset, and strategies were implemented to address data gaps wherever possible. Table 1 presents a summary of the metadata completeness and highlights areas requiring careful consideration during the analysis[5].

TABLE I. COMPLETENESS OF BIBLIOGRAPHIC METADATA

Metadata	Description	Missing Counts	Missing %	Status
AU	Author	0	0.00	Excellent
DT	Document Type	0	0.00	Excellent
LA	Language	0	0.00	Excellent
PY	Publication Year	0	0.00	Excellent
TI	Title	0	0.00	Excellent
TC	Total Citation	0	0.00	Excellent
SO	Journal	13	0.26	Good
AB	Abstract	24	0.48	Good
C1	Affiliation	130	2.61	Good

DI	DOI	439	8.81	Good
DE	Keywords	672	13.48	Acceptable
ID	Keywords Plus	800	16.05	Acceptable
RP	Corresponding Author	1741	34.92	Poor
CR	Cited References	4985	100.00	Completely missing
WC	Science Categories	4985	100.00	Completely missing

3. RESULTS

3.1. Main Information

Our bibliometric analysis, spanning the years 2018 to 2024, encompasses a diverse collection of 4985 documents sourced from 1763 different outlets, including journals, books, and other scholarly publications. The annual growth rate of the dataset reflects a notable -40.94%, suggesting a dynamic landscape with fluctuations in research output over the specified period[6, 7].

Examining the document contents, Figure 1 illustrates the comprehensive nature of our dataset. The average age of documents is 1.83 years, indicative of the timeliness of the research within the field. Each document, on average, is associated with 8.64 citations, highlighting the scholarly impact and relevance of the collected works.

In terms of authorship, our dataset involves the contributions of 14,470 authors. Interestingly, 379 documents are single-authored, shedding light on the individual scholarly endeavors within the domain of trustworthiness in AI research. The collaboration aspect is prominent, with an average of 4.08 co-authors per document, underscoring the collective and interdisciplinary nature of contemporary research in this field. Furthermore, international collaboration is evident, constituting 25.8% of co-authorships, emphasizing the global perspective and collaborative efforts in advancing the understanding of trustworthy AI.

The variety of document types adds nuance to the dataset. While the majority are articles (1891), there is also diversity in formats, including books, book chapters, conference papers, reviews, and editorials. The presence of different document types indicates the multifaceted nature of research contributions, ranging from in-depth explorations in articles to broader perspectives in books and conferences.

In examining the thematic richness of the dataset, keywords play a crucial role. The dataset boasts a substantial number of keywords, with 23,999 from Keywords Plus (ID) and 12,509 from Author's Keywords (DE). This abundance of keywords signifies the breadth and depth of topics covered, contributing to a nuanced understanding of the facets of trustworthiness in AI research. Figure 2 visually represents these key findings, providing a comprehensive overview of the main information derived from our bibliometric analysis.

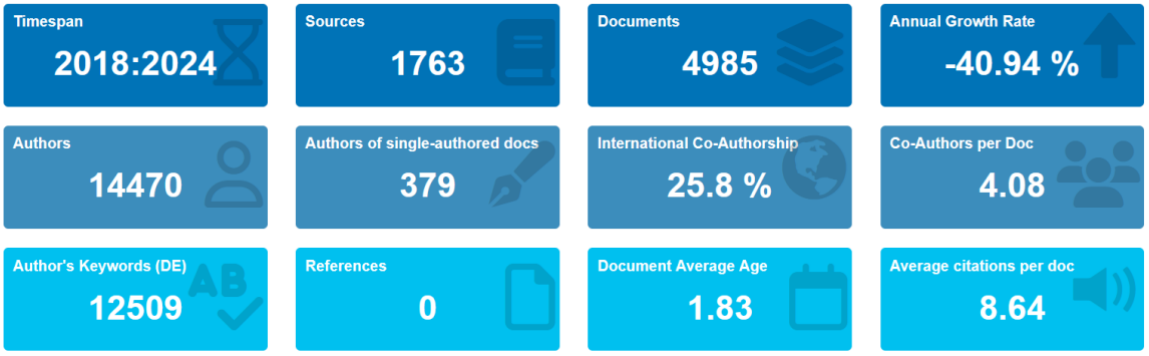

Fig.1. the main information

3.2. Annual Scientific Production

The analysis of annual scientific production, depicted in Figure 3, unveils intriguing patterns in the evolution of research output within the field of trustworthiness in AI. The dataset, spanning from 2018 to 2024, reflects a notable upward trajectory in scholarly contributions[8, 9].

In 2018, the field witnessed an initial surge with 424 articles, laying the groundwork for subsequent years. The following year, 2019, exhibited a 15.7% increase, reaching 491 articles. This upward trend continued in 2020, marking a substantial 43.9% rise with 706 articles. The year 2021 saw a further 23.1% growth, with 871 articles contributing to the scholarly discourse on trustworthiness in AI.

The most remarkable surge occurred in 2022, where the annual scientific production escalated to 1199 articles, signifying a substantial 37.7% increase from the previous year. This surge might be indicative of increased interest, funding, or technological advancements driving research in AI trustworthiness during that period. The momentum sustained into 2023, with 1276 articles, maintaining the field's robust scientific production[10].

It is essential to note that the year 2024 exhibits a seemingly significant decline in output, with only 18 articles. However, this dip may be attributed to the ongoing nature of the publishing process; as of the analysis cutoff in 2023, several publications earmarked for 2024 may still be in the pipeline.

In summary, the annual scientific production trend indicates a dynamic and growing interest in trustworthiness in AI research, with intermittent fluctuations that could be influenced by various internal and external factors. Understanding these patterns provides valuable insights into the evolving landscape of research contributions within this critical domain.

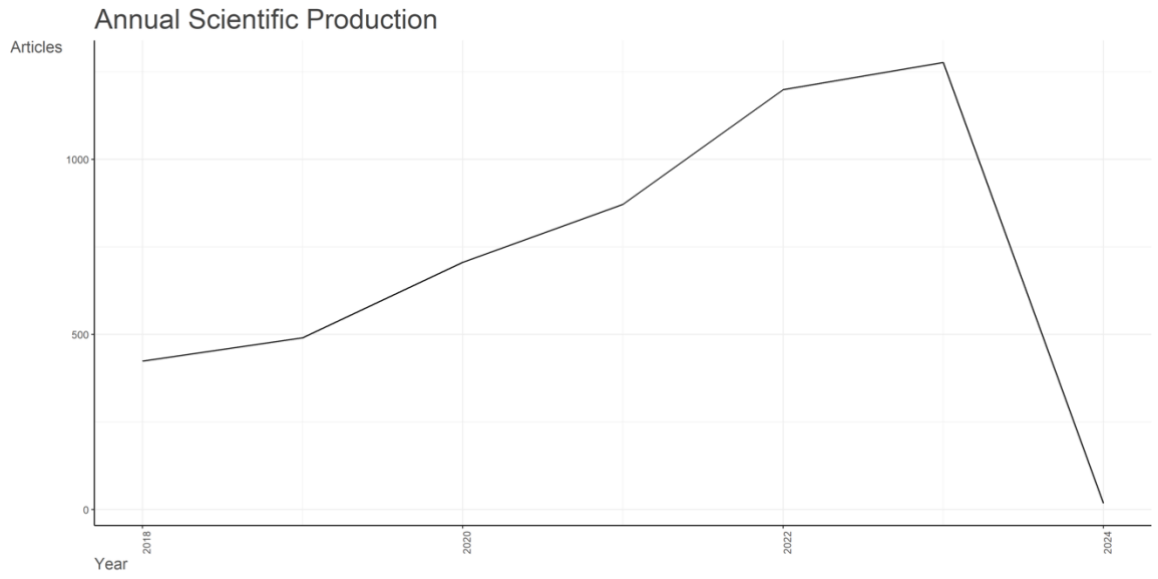

Fig. 3. Annual scientific production

3.3. Average Citations per Year

Figure 4 and Table 2 provides a comprehensive overview of the annual citation trends per year, shedding light on the impact and longevity of scholarly contributions in the realm of trustworthiness in AI. The Mean Total Citations per Article (Mean TC per Art) for each year, calculated by dividing the total citations by the number of articles, serves as a key metric[11, 12].

In 2018, the Mean Total Citations per Article was notably high at 20.36, reflecting a substantial impact of the published articles[13]. However, the Mean Total Citations per Year (Mean TC per Year) declined to 3.39, indicating that, on average, each article received citations over a span of six years[12]. This suggests a sustained interest in and relevance of the publications over an extended period.

The subsequent years, 2019 and 2020, experienced a decline in Mean Total Citations per Article to 13.64 and 13.15, respectively. Although the Mean Total Citations per Year also decreased, hovering around 2.73 and 3.29, the publications from these years maintained a sustained impact, with an average citation duration of five and four years, respectively[14].

The year 2021 exhibited a slight decrease in Mean Total Citations per Article to 12.94, yet the Mean Total Citations per Year increased to 4.31, indicating that the articles from this year garnered citations over an average span of three years. This suggests a notable impact and quicker recognition of the contributions made in 2021[15].

In 2022, the Mean Total Citations per Article dropped significantly to 4.78, and the Mean Total Citations per Year reduced to 2.39. This decrease may suggest a change in the citation dynamics, possibly influenced by shifts in research focus or the emergence of new trends in the field.

The year 2023 saw a substantial decline in both Mean Total Citations per Article (1.13) and Mean Total Citations per Year (1.13). The drastic reduction in Mean Total Citations per Year to match the Mean Total Citations per Article implies a quicker recognition and citation of the works within the same year of publication.

It is essential to approach the data for 2024 cautiously, as the year is ongoing at the time of analysis reflecting that the average citation duration cannot be accurately calculated without complete annual citation data. As the publications for 2024 continue to accumulate citations, a more accurate assessment of their impact can be made in subsequent analyses.

TABLE II. AVERAGE CITATIONS PER YEAR

Year	N	Citable Years	Mean TC per Year	Mean TC per Art
2018	424	6	3.39	20.36
2019	491	5	2.73	13.64
2020	706	4	3.29	13.15
2021	871	3	4.31	12.94
2022	1,199	2	2.39	4.78
2023	1,276	1	1.13	1.13
2024	18	0		0.22

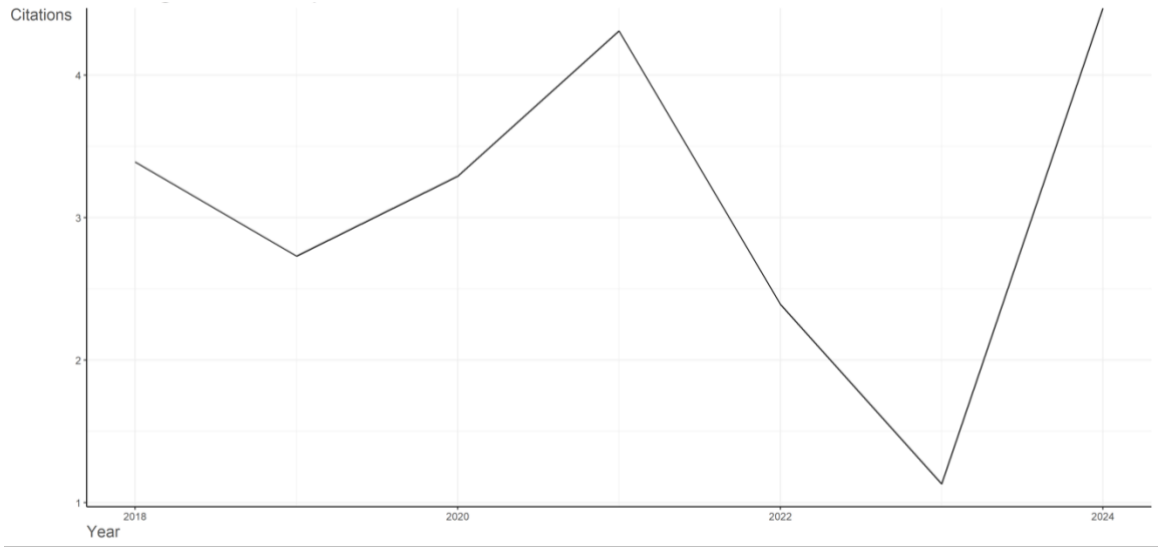

Figure 4: Annual citation trends per year

Most Relevant Sources in Trustworthy AI Research:

Understanding the most relevant sources in the field of trustworthy AI research is crucial for researchers seeking to stay abreast of the latest developments. The analysis of the dataset highlights key outlets where scholarly activity on trustworthiness in AI is concentrated.

The LECTURE NOTES IN COMPUTER SCIENCE (INCLUDING SUBSERIES LECTURE NOTES IN ARTIFICIAL INTELLIGENCE AND LECTURE NOTES IN BIOINFORMATICS), with 271 articles, stands out as the predominant source, emphasizing the significance of conferences and lecture notes in disseminating trustworthy AI research findings. The interdisciplinary nature of this series indicates a broad engagement with AI-related themes.

Following closely is IEEE ACCESS, known for its open-access model, with 162 articles. It demonstrates its influence in providing a platform for a diverse range of AI research. The substantial number of articles indicates the significance of this journal in facilitating accessible and impactful contributions to the trustworthy AI discourse.

The ACM INTERNATIONAL CONFERENCE PROCEEDING SERIES, contributing 140 articles, is a prominent source, reflecting the intersection of computer science and AI. The volume of articles suggests that conferences play a pivotal role in shaping the discourse on trustworthy AI, fostering collaboration and knowledge exchange.

Another significant source is the CEUR WORKSHOP PROCEEDINGS with 120 articles. This indicates the importance of workshops in fostering dialogue and exploration within the trustworthy AI community. The considerable number of articles underscores the role of specialized forums in this research domain.

Finally, the ADVANCES IN INTELLIGENT SYSTEMS AND COMPUTING series, with 105 articles, serves as a platform for advancing knowledge in intelligent systems, playing a significant role in the dissemination of trustworthy AI research. The diverse range of topics covered in this series reflects the multidisciplinary nature of the field[16].
As Figure 5 illustrates, these sources collectively contribute to the dynamic landscape of trustworthy AI research. Researchers can leverage insights from these key outlets to identify the most influential platforms shaping the discourse, explore emerging trends, and engage with the broader scholarly community in this evolving field.

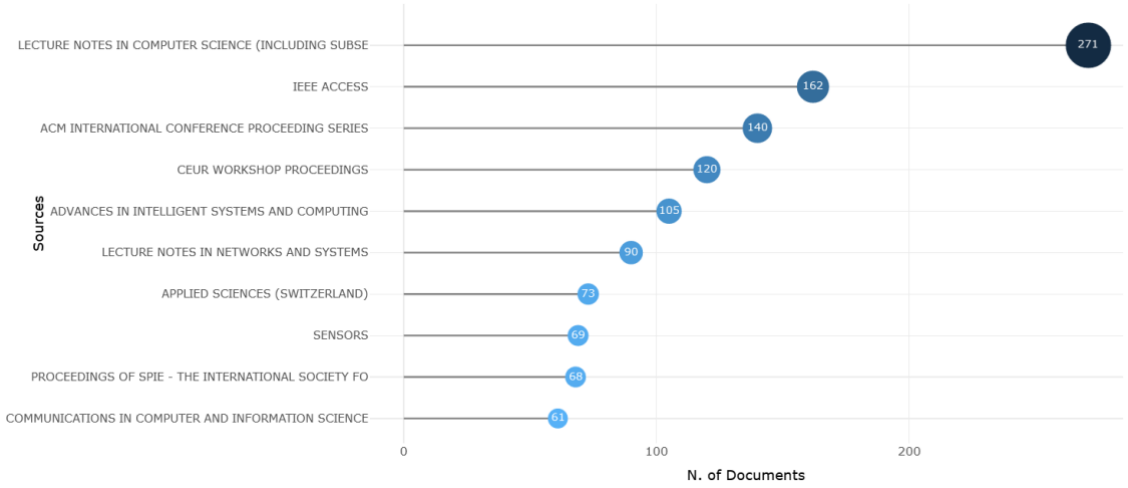

Fig. 5. Most Relevant Sources in Trustworthy AI Research

3.4. Core Sources by Bradford's Law in Trustworthy AI Research

Applying Bradford's Law, as encapsulated by Equation (1), which anticipates the distribution of scientific productivity across journals, facilitates the recognition of core sources that substantially contribute to scholarly output. The examination of the dataset unveils a discernible pattern in the distribution of articles across journals, emphasizing the concentration of impactful sources. This mathematical model aids in understanding the structure of information dissemination within the analyzed field, guiding the identification of pivotal sources that play a crucial role in shaping the scholarly landscape[17].

$$B_n = B_1 \, r^{n-1} \tag{1}$$

Where
(B_n) is the number of journals containing n articles
(B_1) is the number of journals containing one article
(r) Is the Bradford ratio, a constant indicating the rate at which the number of journals containing articles decreases as the number of articles increases.
At the core of trustworthy AI research, the LECTURE NOTES IN COMPUTER SCIENCE (INCLUDING SUBSERIES LECTURE NOTES IN ARTIFICIAL INTELLIGENCE AND LECTURE NOTES IN BIOINFORMATICS) emerges as a pivotal source, consistent with its prominence in the most relevant sources section. According to Bradford's Law, this series serves as a primary source, hosting a substantial portion of articles, demonstrating its central role in disseminating research findings in the trustworthy AI domain.
Following the core source, IEEE ACCESS aligns with its significant representation in the most relevant sources, reaffirming its importance in the field. The Law predicts that IEEE ACCESS plays a crucial role as a secondary source in the dissemination of trustworthy AI research.
The ACM INTERNATIONAL CONFERENCE PROCEEDING SERIES also exhibits a concentrated presence in line with Bradford's Law, further corroborating its significance in scholarly communication within the trustworthy AI community. The Law suggests that this series functions as a tertiary source, emphasizing its role in complementing the core outlets.
As Figure 6 visually demonstrates, the distribution of articles across these core sources aligns with the predicted pattern outlined by Bradford's Law. This analytical approach provides valuable insights into the hierarchical structure of sources in trustworthy AI research. Researchers can leverage this understanding to efficiently navigate the scholarly landscape, focusing on core outlets to stay informed about pivotal developments in the field.

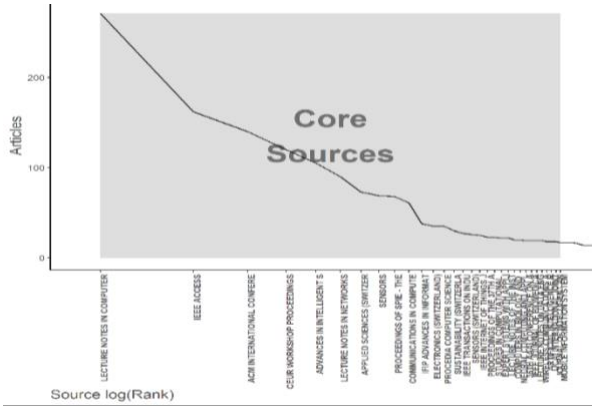

Fig. 7. scientific productivity across journals using Bradford's Law

3.5. Sources' Local Impact in Trustworthy AI Research

An essential aspect of evaluating the scholarly impact of sources in the field of trustworthy AI research involves examining various bibliometric indicators. The elements such as h-index, g-index, and m-index provide insights into the sources' local influence within the research community. Figure 8 visually represents the impact of selected sources, allowing for a comparative analysis of their performance.

Among the sources under consideration, IEEE ACCESS stands out with an h-index of 23, indicating that 23 articles from this source have been cited at least 23 times each. This substantial h-index is complemented by a g-index of 48, which considers not only highly cited articles but also the total number of citations they receive. Furthermore, the m-index of 3.833 suggests a consistent impact, considering the median number of citations received by the articles.

Similarly, APPLIED SCIENCES (SWITZERLAND) and LECTURE NOTES IN COMPUTER SCIENCE (INCLUDING SUBSERIES LECTURE NOTES IN ARTIFICIAL INTELLIGENCE AND LECTURE NOTES IN BIOINFORMATICS) exhibit notable local impact, with comparable h-indices of 13. The g-index and m-index further support their significance, demonstrating a balanced influence across their respective articles.

SENSORS and SENSORS (SWITZERLAND), both integral sources in trustworthy AI research, present unique characteristics. While SENSORS showcases a higher h-index (13) and m-index (4.333), indicating a higher median impact, SENSORS (SWITZERLAND) has a slightly higher g-index (26), emphasizing its cumulative impact[18].

Examining journals like IEEE INTERNET OF THINGS JOURNAL and IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, with h-indices of 12, reinforces their consistent local impact in the field. These sources exhibit steady performance across various bibliometric measures, contributing significantly to the scholarly landscape.

ELECTRONICS (SWITZERLAND), NEURAL COMPUTING AND APPLICATIONS, and SUSTAINABILITY (SWITZERLAND) also play crucial roles, each showcasing specific strengths in terms of local impact. These sources contribute to the diverse scholarly ecosystem of trustworthy AI research, with varying h-indices and citation patterns.

Understanding the local impact of sources is vital for researchers, allowing them to gauge the influence of specific journals on the field. While the h-index emphasizes citation quantity, the g-index considers highly cited articles, and the m-index provides insights into the median impact. This multifaceted analysis aids researchers in making informed decisions about where to publish and where to find high-impact content in the realm of trustworthy AI.

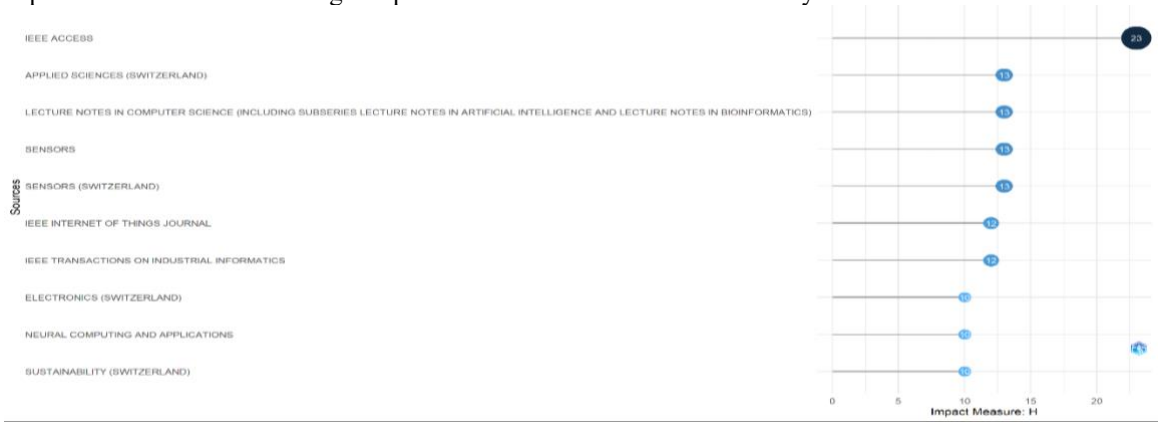

Fig. 8. Source's local impact

3.6. Sources' Production over Time in Trustworthy AI Research

Analyzing the temporal evolution of sources' production provides valuable insights into the dynamics of scholarly contributions in the field of trustworthy AI research. Figure 9 visually represents the production trends of selected sources, showcasing their annual output from 2018 to 2024[19].

LECTURE NOTES IN COMPUTER SCIENCE (LNCS), IEEE ACCESS, ACM INTERNATIONAL CONFERENCE PROCEEDING SERIES, CEUR WORKSHOP PROCEEDINGS, and ADVANCES IN INTELLIGENT SYSTEMS AND COMPUTING are prominent sources contributing significantly to the scholarly output in this domain.

The year 2018 marked a modest beginning, with LNCS producing 42 articles, IEEE ACCESS contributing 9, ACM International Conference Proceeding Series with 24, CEUR Workshop Proceedings publishing 7, and Advances in Intelligent Systems and Computing presenting 23 articles.

Subsequent years witnessed a substantial increase in production across all selected sources. In 2019, the production for LNCS surged to 98 articles, IEEE ACCESS to 21, ACM International Conference Proceeding Series to 40, CEUR Workshop Proceedings to 18, and Advances in Intelligent Systems and Computing to 42.

The upward trajectory continued in 2020, with LNCS reaching 140 articles, IEEE ACCESS contributing 37, ACM International Conference Proceeding Series publishing 77, CEUR Workshop Proceedings presenting 40, and Advances in Intelligent Systems and computing showcasing 79 articles.

The year 2021 witnessed a notable increase in production across the board, with LNCS producing 184 articles, IEEE ACCESS contributing 63, ACM International Conference Proceeding Series publishing 106, CEUR Workshop Proceedings presenting 59, and Advances in Intelligent Systems and Computing showcasing 105 articles[20].

In 2022 and 2023, the sources maintained a steady production rate. LNCS, IEEE ACCESS, ACM International Conference Proceeding Series, CEUR Workshop Proceedings, and Advances in Intelligent Systems and computing consistently contributed to the scholarly output in trustworthy AI research[21].

While 2024 shows consistent numbers with 2023, it is essential to consider that the year is still ongoing, and the data may not be complete. Nonetheless, the overall trend indicates a robust and growing scholarly production, emphasizing the sustained efforts and contributions of these key sources to the evolving landscape of trustworthy AI research over the analyzed period.

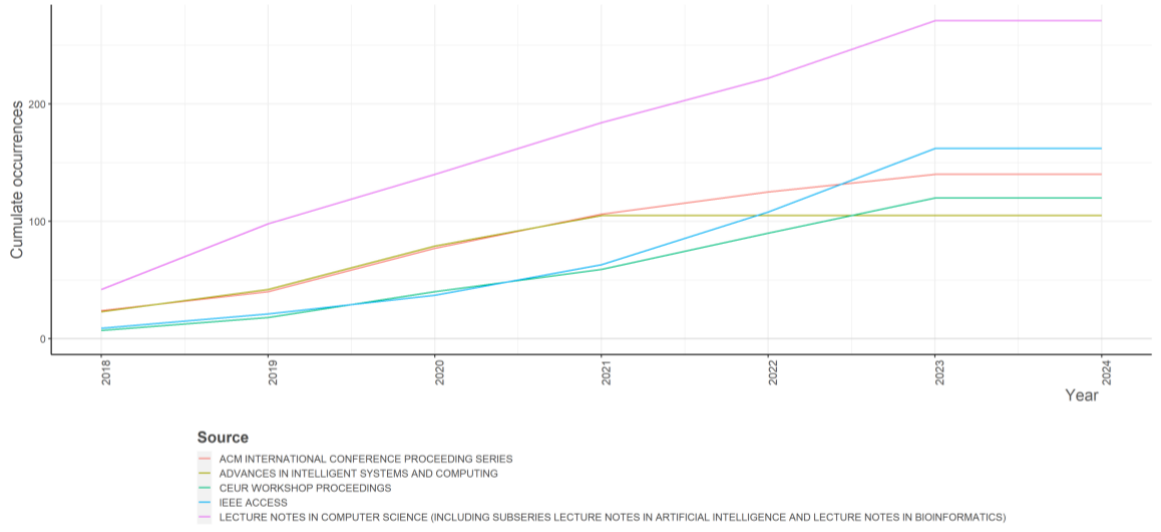

Fig. 9. Sources' Production over Time

3.7. Most Relevant Authors

Figure 10 presents a comprehensive overview of the most relevant authors actively contributing to the discourse on trustworthy AI, based on the analysis of their publications from 2018 to 2024. The authors' impact is assessed not only by the total number of articles authored but also by considering fractionalized contributions, providing insights into their consistent presence and influence across various publications.

Wang Y emerges as a leading figure in the field, having authored a substantial 70 articles. This prolific output represents a significant 18.93% of the total scholarly contributions, indicating Wang Y's profound and sustained engagement with the

subject matter. The high fractionalized value underscores not just the quantity but also the consistent quality and impact of their work across a diverse range of publications.

Closely following is Liu Y with 56 articles, constituting a notable 12.31% of the overall contributions. Liu Y's substantial presence in the field suggests a consistent and valuable contribution to the literature on trustworthy AI.

Li Y, with 54 articles (12.00%), and Zhang J with 51 articles (11.86%), occupy key positions in the list of most relevant authors. These scholars have demonstrated a sustained and impactful presence in the scholarly discourse surrounding trustworthy AI.

The list further includes Zhang Y (11.38%), Zhang Z (11.22%), Li X (11.36%), Wang Z (10.60%), and Zhang X (10.75%). Each of these authors has made substantial contributions, both individually and collectively shaping the landscape of trustworthy AI research.

The fractionalized contributions provide a nuanced understanding of the authors' impact, emphasizing not just the quantity of publications but also their consistent engagement across a diverse range of research avenues. This analysis serves as a valuable guide for researchers, policymakers, and practitioners seeking authoritative voices in the evolving domain of trustworthy AI.

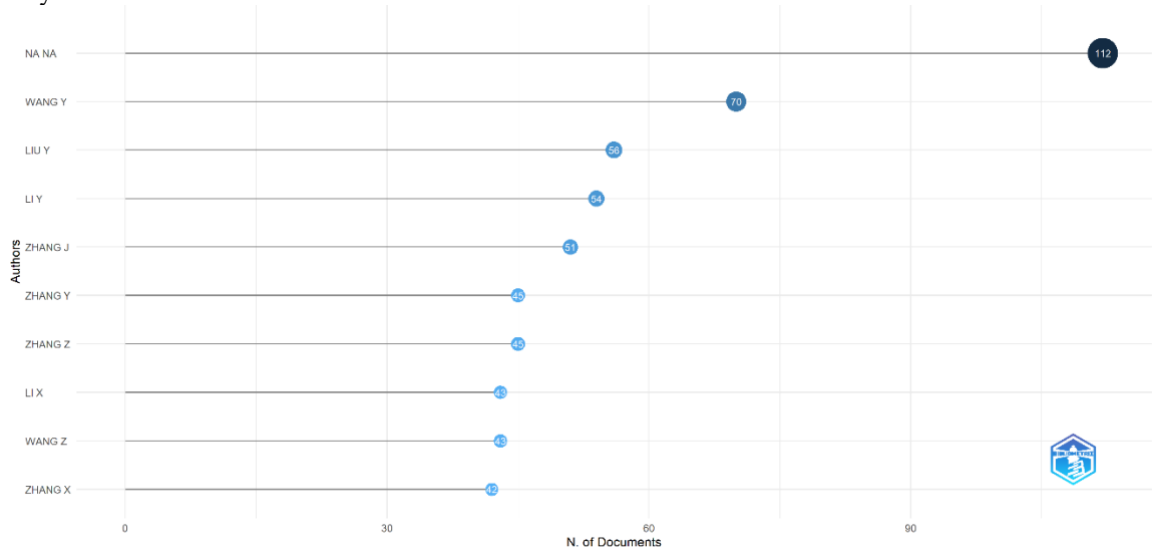

Fig. 10. Most Relevant Authors

3.8. Authors' Production Over Time

Figure 11 provides an insightful perspective on the temporal evolution of authors' scholarly production in the realm of trustworthy AI research. Analyzing the data from 2018 to 2024 reveals nuanced patterns, shedding light on prolific contributors and the trajectory of their contributions over the years.

Li X, a prominent figure in the field, exhibits a steady increase in productivity from 2018 to 2023, peaking at 11 articles. This ascent highlights sustained engagement, with a temporary dip in 2020. The Total Citations per Year (TCpY) metric underscores the impact of Li X's work, with 97 citations in 2022, indicating the enduring relevance of their contributions.

Similarly, Li Y demonstrates a robust presence with a peak of 12 articles in 2019, coupled with a significant TCpY of 79.2 in the same year. Despite a decline in article frequency in subsequent years, Li Y maintains a notable impact, as evident from the citations garnered.

Liu Y showcases a dynamic trajectory, with a substantial increase in output over the years, reaching a peak of 14 articles in 2023. The TCpY of 39 in 2023 accentuates the growing influence of Liu Y's work, indicating an upward trajectory in both quantity and impact.

Wang Y exhibits consistent productivity, with a peak of 18 articles in 2023, although 2024 shows a decline in output. The TCpY metric fluctuates but demonstrates a notable impact, particularly in 2020 with 115 citations.

Wang Z, Zhang J, and Zhang X each showcase distinctive patterns. Wang Z and Zhang J experience peaks in productivity, and their TCpY metrics reflect impactful contributions, while Zhang X's output diminishes after 2021.

This analysis provides a comprehensive understanding of authors' production trends, illustrating variations in output and impact over the specified timeframe. The TCpY metric serves as a valuable indicator, emphasizing the enduring influence of certain authors within the evolving landscape of trustworthy AI research.

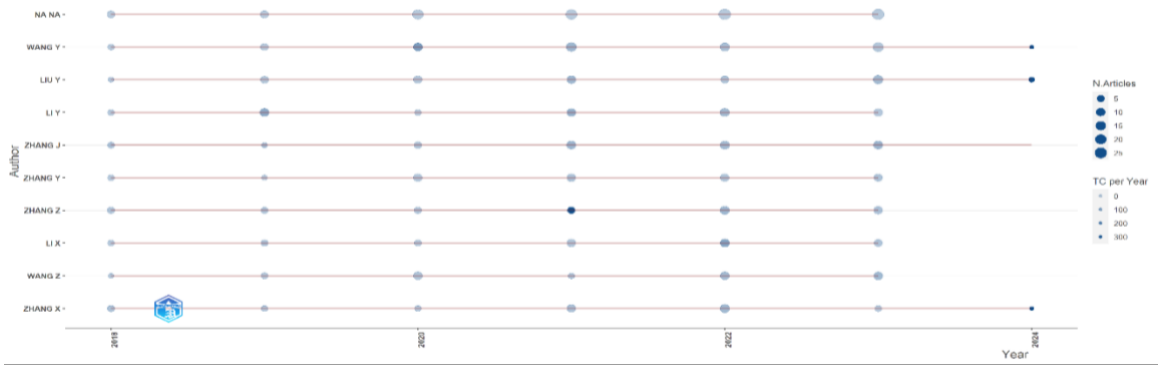

Fig. 11. Authors' Production over Time

3.9. Author Productivity Through Lotka's Law

Lotka's Law, a fundamental principle in bibliometrics, is represented by Equation (2) below, offering valuable insights into the distribution of productivity among authors. The examination of trustworthy AI research, depicted in Figure 12, aligns with Lotka's Law, highlighting a discernible pattern in the distribution of scholarly output among authors. This mathematical model helps us understand the concentration of authorship and the unequal distribution of research productivity within the analyzed dataset.

$$\frac{1}{n} = \frac{1}{a} \times (C)^{-b} \tag{2}$$

Where:
(n) Is the number of authors who have written x papers,
(a) Is a constant.
(C) Is the total number of authors,
(b) Is a constant.
The data presents a breakdown of the number of documents written by individual authors, showcasing the distribution of productivity. Notably, a substantial proportion of authors, approximately 85%, have authored only one document, reflecting a prevalent trend where a large number of contributors engage in singular scholarly endeavors within the specified timeframe.
As the number of documents per author decreases, the corresponding proportion of authors increases, following a power-law distribution characteristic of Lotka's Law. The law suggests that a small percentage of authors contribute a significant portion of the total scholarly output, while the majority of authors produce fewer documents.
Examining the graph in Figure 12, it becomes evident that the distribution aligns with Lotka's Law, emphasizing the concentration of productivity among a select group of authors. This observation underscores the presence of prolific contributors who significantly impact the field of trustworthy AI research.
While the law accurately captures the distribution pattern, it is essential to recognize the diverse contributions within this ecosystem. The presence of authors with multiple documents, though a minority, signifies the existence of influential contributors whose work substantially contributes to the scholarly landscape. Lotka's Law provides a valuable lens through which to understand the dynamics of author productivity in trustworthy AI research. The concentration of output among a limited number of authors highlights the importance of recognizing and acknowledging the significant contributions made by these prolific scholars.

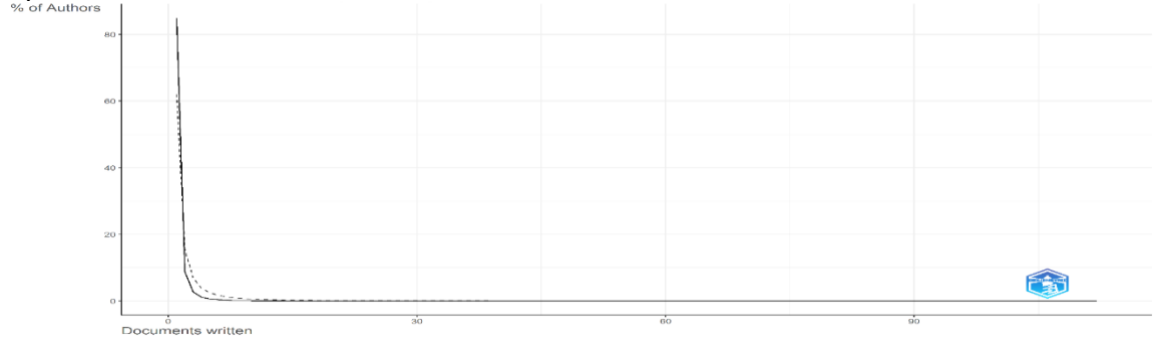

Fig. 12. Author Productivity Through Lotka's Law

3.10. Authors Local Impact

The examination of local impact metrics for authors, particularly those with a Hirsch index (h-index) of 9 and over, sheds light on the influence and significance of their scholarly contributions within the trustworthy AI research domain. Figure 13 visually represents the key indicators, while the accompanying table outlines the detailed metrics for each author.

Among the highlighted authors, Li J stands out with a notable h-index of 12, Girsch index (g-index) of 24, and m-quotient (m-index) of 2, reflecting a substantial impact on the field. Since the start of their scholarly contributions in 2018, Li J has amassed a Total Citations (TC) count of 622, indicating the reach and influence of their work. Similarly, Zhang J, with an h-index of 12, has made a considerable impact, accumulating 582 citations since 2018.

Li Y and Liu Y, both with an h-index of 11, demonstrate consistent scholarly impact, reflected in their Total Citations counts of 706 and 322, respectively. Zhang Z, with an h-index of 11, has a remarkable Total Citations count of 1259, showcasing significant influence within the field since 2018.

Authors with an h-index of 10, including Liu X, Zhang X, Zhang Y, and Li X, exhibit substantial local impact, each contributing uniquely to the scholarly landscape. Their contributions are characterized by consistent citation rates and influential publications.

Noteworthy contributors with an h-index of 9, such as Wang J, showcase a commendable Total Citations count of 1977, underlining their sustained influence within the trustworthy AI research community. The local impact analysis underscores the significance of these authors' contributions and their role in shaping the scholarly discourse in the field.

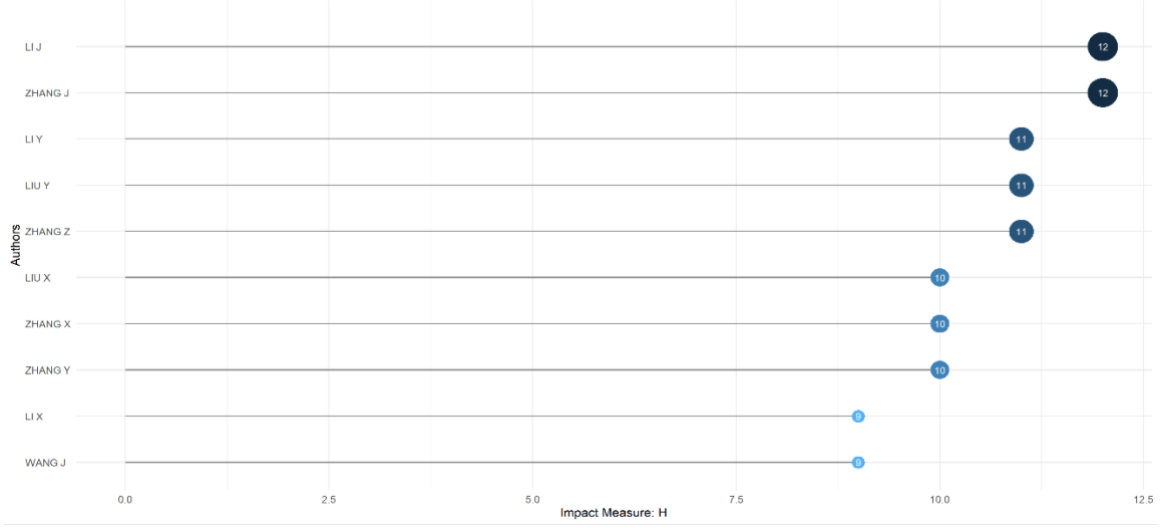

Fig. 13. Local impact metrics for authors

3.11. Most Relevant Affiliations

The examination of the most relevant affiliations within the realm of trustworthy AI research provides valuable insights into the institutional landscape shaping this field. Figure 14 visually represents the distribution of scholarly output among these affiliations, while the accompanying table details the number of articles associated with each institution.

Leading the list is Tsinghua University, with a substantial contribution of 48 articles, showcasing its prominent role in fostering research on trustworthy AI. Following closely is Beijing University of Posts and Telecommunications, which has demonstrated a strong commitment to this area with 43 associated articles.

A noteworthy observation is the presence of affiliations labeled as "Not Reported," indicating a degree of ambiguity regarding the reporting of institutional affiliations in the dataset. Nevertheless, these entries contribute significantly to the overall scholarly output in the field, with 40 associated articles.

The University of California, with 33 articles, stands out as a key contributor to trustworthy AI research, reflecting the diverse and impactful contributions from this institution. Similarly, Huazhong University of Science and Technology and Beihang University, with 32 and 30 articles, respectively, play integral roles in advancing research within the trustworthy AI domain.

The University of Chinese Academy of Sciences, University of Pisa, Zhejiang University, and Northeastern University are additional institutions that have made noteworthy contributions, each having 30 or more associated articles. These findings underscore the global nature of trustworthy AI research, with institutions from various regions actively engaging in scholarly endeavors to advance the field.

The analysis of the most relevant affiliations emphasizes the collaborative and diverse nature of research on trustworthy AI, with a multitude of institutions contributing to the development and dissemination of knowledge in this critical domain.

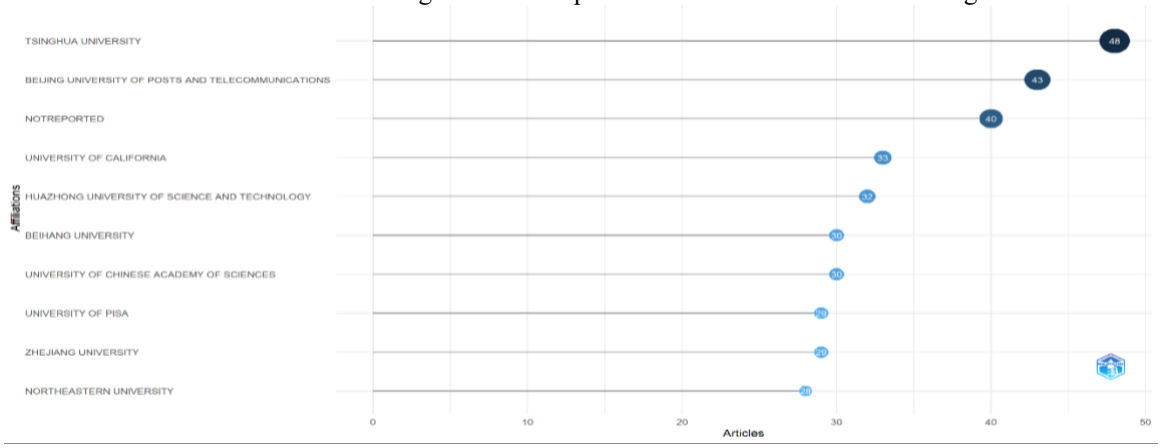

Fig. 14. most relevant affiliations

3.12. Corresponding Authors Countries

The study, on authorship distribution in AI research provides insights into how different countries contribute to efforts in this field. Figure 15 visually presents the distribution while the accompanying table offers information about the number of articles associated with each country.

The dataset consists of a total of 1794 articles, out of which 1450 have a Single Corresponding Author (SCP) and 344 feature Multiple Corresponding Authors (MCP). The country of the corresponding author plays a role in understanding the distribution of research contributions. China emerges as a leading contributor with 833 articles showcasing its presence in AI research. India follows closely with 319 articles indicating its growing contribution to the field. The USA demonstrates research output with 268 articles highlighting its enduring prominence in AI related endeavors. Korea, Italy and Germany also make contributions to the scholarly landscape each having over 100 articles. The United Kingdom and Spain exhibit research output with 115 and 95 articles respectively. These findings highlight the collaboration and participation in reliable AI research.

Analyzing the MCP ratio, which represents the frequency of articles, with corresponding authors, uncovers patterns. Countries such, as the United Kingdom and Spain exhibit MCP ratios, which implies a tendency towards research practices in these regions. Australia, Saudi Arabia and Malaysia also exhibit characteristics with lower article counts but higher MCP ratios. This suggests that while these countries may have contributions their researchers are more inclined towards collaborative authorship. Analyzing authorship provides insights into the geographic distribution of trustworthy AI research contributions. The prominence of efforts as indicated by the MCP ratio further underscores the cooperative nature of research endeavors, in this dynamic field.

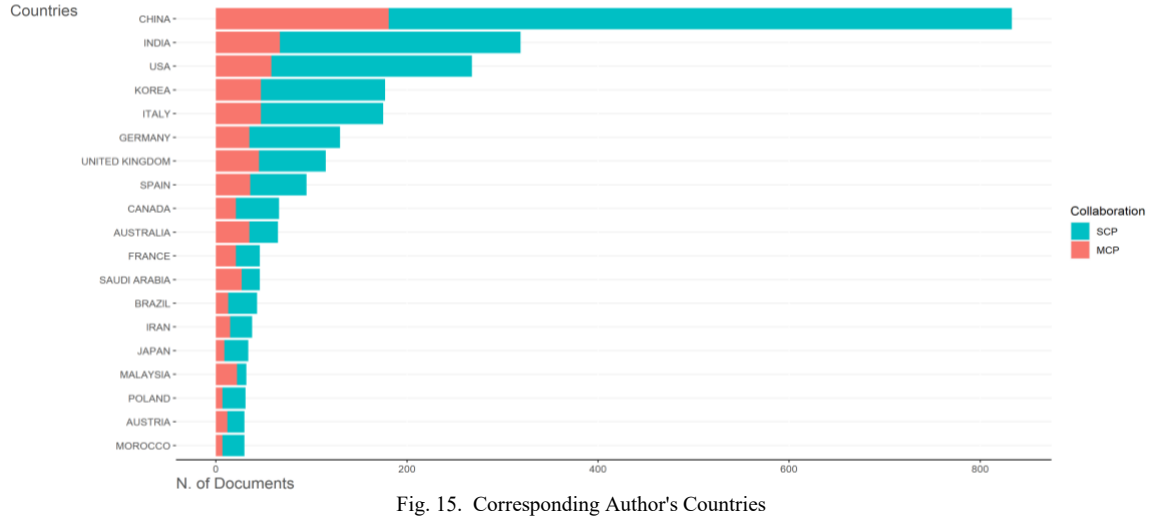

Fig. 15. Corresponding Author's Countries

3.13. Global Dynamics of Trustworthy AI Research: Production, Trends, and Impact

The global landscape of trustworthy AI research is characterized by a rich tapestry of contributions from countries around the world, as evident in the distribution of scientific production represented in Figure 16. China emerges as a prominent contributor, leading with 2414 articles, followed closely by the USA with 1276 articles. Other notable contributors include India, Italy, and Germany, reflecting the diverse international engagement in trustworthy AI research.

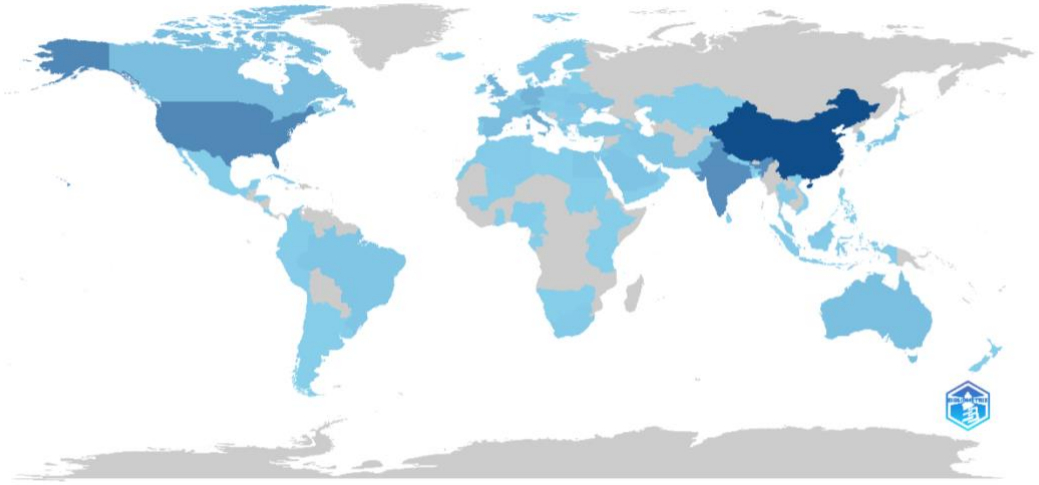

Fig. 16. Global Distribution of Trustworthy AI Research

Figure 17 provides a temporal perspective on countries' production over time, illustrating a consistent upward trajectory for China, which reached a peak of 2414 articles in 2024. Similarly, Germany and India exhibit steady growth, underscoring their sustained commitment to trustworthy AI research.

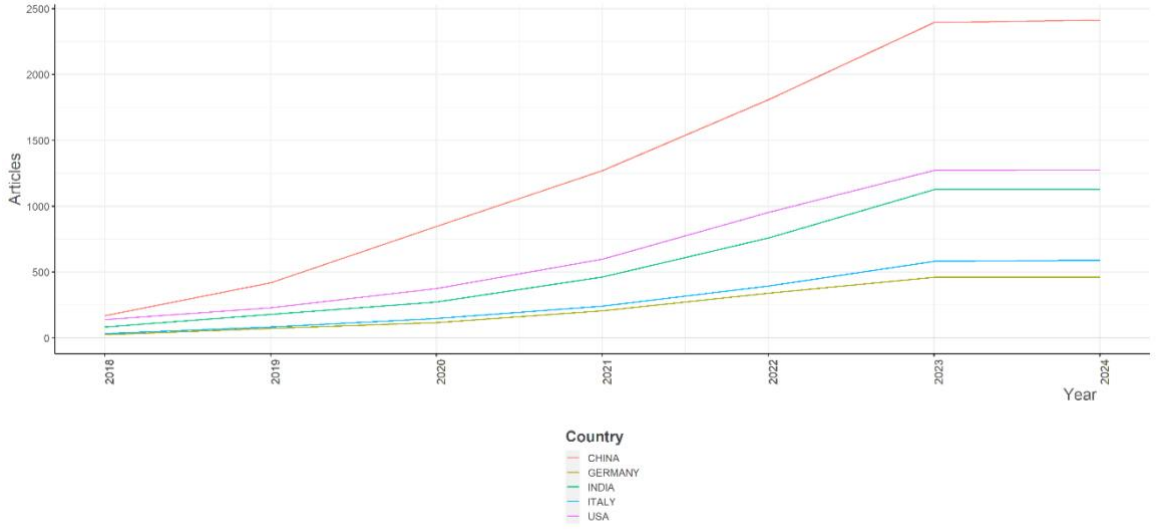

Fig. 17. Temporal Evolution of AI Research Production by Country

Turning attention to Figure 18, the most cited countries, a correlation between high scientific production and citation impact becomes apparent. China, with 9207 total citations, maintains an average article citation of 11.10. The USA, with 3970 total citations, demonstrates a higher average article citation of 14.80, indicative of the influential nature of its research.

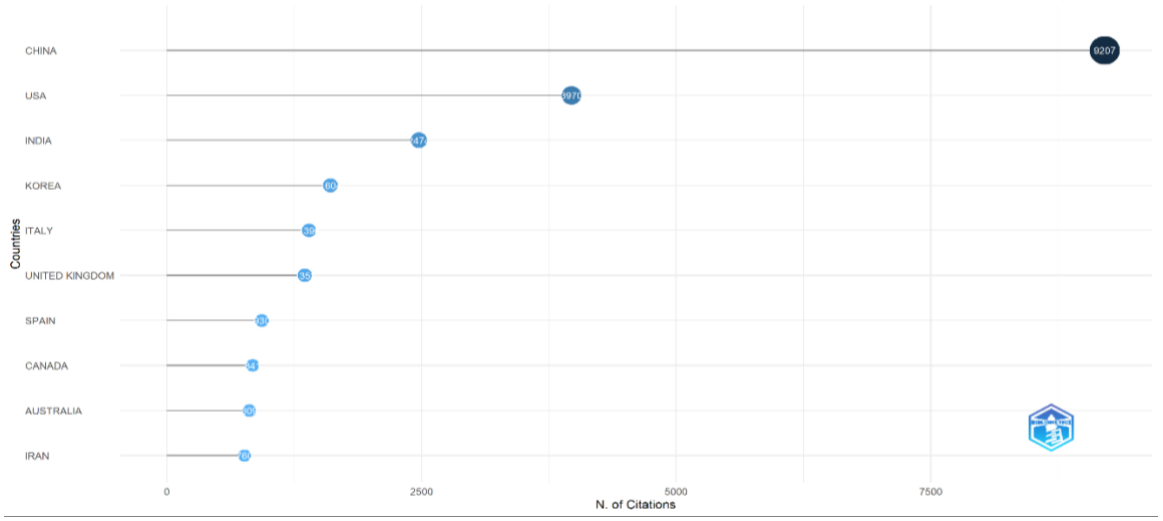

Fig. 18. Most Cited Countries in Trustworthy AI Research

Countries like India, Korea, and Italy also exhibit a correlation between research impact and citation frequency, with average article citations ranging from 7.80 to 9.10. Noteworthy outliers include Hong Kong, Singapore, and Denmark, showcasing exceptionally high average article citations and highlighting the influential contributions of these nations in the field.

In summary, the interaction between countries' scientific production, production trends over time, and citation impact offers a comprehensive understanding of the global landscape of trustworthy AI research. This collaborative effort across diverse nations contributes to the advancement and impact of this dynamic field on a global scale.

3.14. Most Global Cited Documents

In the realm of Trustworthy AI research, certain documents have emerged as global benchmarks, accumulating significant citations and shaping the discourse in the field. Table 3 provides a comprehensive overview of the Most Global Cited Documents, showcasing their Total Citations (TC), Normalized TC, and TC per Year.

The document authored by Liu et al. in 2018, titled "Mechanical Systems and Signal Processing," stands out prominently with an impressive Total Citations of 1325, contributing significantly to the global discourse on Trustworthy AI. Similarly, the work of You et al. in 2021, "Science China Information Sciences," has garnered substantial attention, boasting a remarkable 274 Normalized TC, indicative of its influential impact in the field.

Tjoa et al.'s 2021 paper, "IEEE Transactions on Neural Networks and Learning Systems," is another noteworthy contribution, with 507 Total Citations and a Normalized TC of 39.17. These documents, along with others listed in Table 3, represent pivotal resources that have shaped and advanced the discussions on Trustworthy AI.

It is crucial to note the diversity in the types of documents that have achieved high citation rates, ranging from journal articles like "Medical Image Analysis" by Signoroni et al. in 2021 to "Annual Review of Control, Robotics, and Autonomous Systems" by Schwarting et al. in 2018. This diversity underscores the multidisciplinary nature of Trustworthy AI research and the various avenues through which impactful contributions are made.

The Most Global Cited Documents, as outlined in Table 3, not only serve as valuable resources for researchers but also reflect the dynamic landscape of Trustworthy AI, where innovation and impactful contributions emanate from a diverse array of studies.

TABLE III. MOST GLOBAL CITED DOCUMENTS

Paper	Total Citations	Normalized TC	TC per Year
LIU R, 2018, MECH SYST SIGNAL PROCESS	1325	65.08	220.83
YOU X, 2021, SCI CHINA INF SCI	822	63.5	274
TJOA E, 2021, IEEE TRANS NEURAL NETWORKS LEARN SYS	507	39.17	169
SCHWARTING W, 2018, ANNU REV CONTROL ROBOT AUTON SYST	497	24.41	82.83
BEDI G, 2018, IEEE INTERNET THINGS J	458	22.49	76.33
SIGNORONI A, 2021, MED IMAGE ANAL	452	34.92	150.67
VISWANATHAN H, 2020, IEEE ACCESS	380	28.9	95
GEETHARAMANI G, 2019, COMPUT ELECTR ENG	368	26.99	73.6
XU H, 2018, IEEE ACCESS	357	17.53	59.5
ZHU G, 2020, IEEE TRANS WIRELESS COMMUN	340	25.86	85

3.15. Authors collaboration network

Collaboration is a cornerstone of academic research, fostering the exchange of ideas and the development of innovative solutions. Figure 19 offers a visual representation of the Collaboration Network, illustrating the interconnected relationships among authors based on various metrics, including Cluster, Betweenness, Closeness, and PageRank.

In the network, each node represents an author, and the clusters denote distinct groups of authors who have demonstrated a high level of collaboration. The size and position of each node reflect its prominence and influence within the network. Notably, authors such as Zhang Z, Wang J, and Wang H are central figures in Cluster 1, characterized by their high Betweenness and PageRank values, indicating their pivotal roles in connecting different parts of the collaboration network.

Cluster 1 comprises authors who exhibit strong ties and frequent collaboration, as evident from the Betweenness values, which signify the extent to which an author connects disparate clusters within the network. The Closeness values, representing how quickly an author can reach other collaborators, highlight the efficiency of information flow within this closely-knit cluster.

Moving to Cluster 2, Wang Y, Li X, and Zhang X emerge as key contributors. Their high PageRank values suggest a substantial influence on the overall network structure. Cluster 2, while distinct, maintains connectivity with Cluster 1, demonstrating a broader collaboration landscape.

In Cluster 3, authors like Liu Y, Li Y, and Zhang J play essential roles in facilitating collaboration across different subgroups. Their high Betweenness values indicate their significance in connecting disparate clusters and fostering a more integrated collaboration network.

Clusters 4, 5, and 6, represented by Chen J, Kim J, and Lee J, exhibit unique patterns of collaboration. Chen J stands out with a notably high Betweenness value, suggesting a bridge-like role in connecting diverse clusters.

It is important to note that while some authors exhibit strong collaboration within their clusters, others, such as Kim J and Lee J, have relatively lower collaborative activity, indicated by their lower Betweenness and PageRank values.

The Collaboration Network depicted in Figure 19 provides a nuanced view of the collaborative dynamics among authors in the realm of Trustworthy AI research. The varying sizes, positions, and interconnections of nodes underscore the complex and multifaceted nature of collaboration in this field.

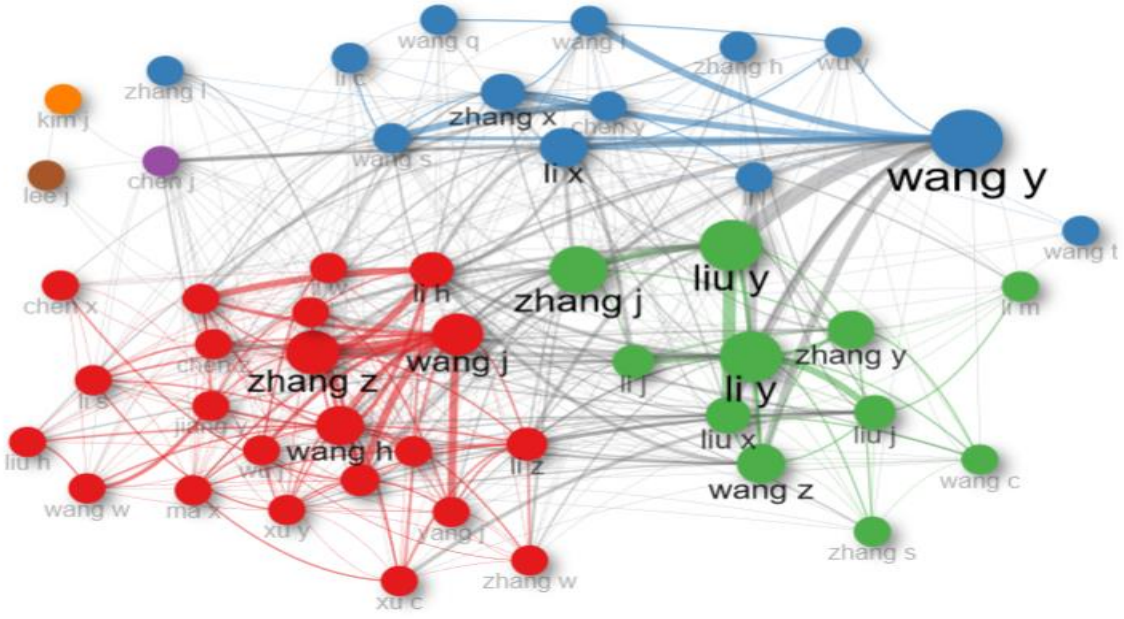

Fig. 19. Authors collaboration network

4. DISCUSSION

In unraveling the intricate tapestry of Trustworthy AI research, our exploration of various facets—ranging from author productivity to collaboration networks—has provided valuable insights into the dynamics of this evolving field. The application of bibliometric principles, such as Lotka's Law and Bradford's Law, has allowed us to discern patterns, trends, and concentrations that bear significance for both current practitioners and future researchers.

Our journey began with an examination of Author Productivity through Lotka's Law. By delving into the distribution of scholarly output among authors, as depicted in Figure 12, we discerned a discernible pattern aligning with Lotka's Law. This distribution underscores the prevalence of prolific authors and the rarity of those with lower productivity. The disproportionate distribution, evident from the steep decline in the number of authors as documents increase, prompts contemplation on the underlying factors influencing such trends. Are there systemic issues limiting broad authorship, or does this pattern reflect the inherent nature of scholarly contributions in Trustworthy AI research?

Moving on to Authors' Local Impact, where the focus was on individuals with an h-index of 9 and over, a nuanced understanding of author impact emerged. The h-index, g-index, m-index, total citations, number of publications, and the start year were dissected to reveal varying degrees of influence and productivity. Authors like LI J and ZHANG J exhibited high impact, backed by substantial total citations and a balanced h-index. This prompts a deeper exploration into the methodologies and topics that contribute to enduring impact, providing guidance for aspiring researchers in the field.

As we explored Most Relevant Affiliations, the critical role of institutional affiliations in shaping Trustworthy AI research became evident. Figure 14, showcasing the distribution of articles across various affiliations, prompts questions about the impact of institutional environments on research outcomes. Does the concentration of articles within certain affiliations indicate a collaborative stronghold or reveal institutional preferences for specific research themes?

Corresponding Author's Countries brought a global perspective to our analysis. The distribution of articles across different countries, as illustrated in Figure 15, emphasizes the global nature of Trustworthy AI research. Notably, China emerges as a dominant force, raising questions about the role of geopolitics, funding, and institutional support in shaping the landscape of scientific production. Is China's prominence a result of concerted national initiatives or reflective of broader global trends?

Shifting our gaze to Countries' Scientific Production, illustrated in Figure 16, we observed a diverse landscape with China, the USA, and India leading the pack. The distribution of scientific production, shown through a world map, prompts contemplation on the factors driving regional variations. What geopolitical, economic, or cultural factors contribute to the disparate scientific output across nations?

Examining Countries' Production over Time, as depicted in Figure 17, unraveled temporal trends in select countries. The significant surge in China's production over the years stands out, posing questions about the factors driving this exponential growth. How do policy changes, technological advancements, and global collaborations contribute to the temporal evolution of scientific production in Trustworthy AI research?

The Most Cited Countries section spotlighted the influence and impact of countries based on average article citations. Figure 18 illustrates the interplay between scientific production and citation impact. Notably, countries like Hong Kong, Singapore, and Belgium demonstrate high citation impact despite a comparatively lower scientific output. This raises intriguing questions about the qualitative aspects of research emanating from these regions.

Transitioning to the realm of Most Global Cited Documents, Table 3 introduces us to seminal works that have left an indelible mark on Trustworthy AI research. These documents, distinguished by their total citations, normalized TC, and paper titles, serve as the bedrock for future investigations. What makes these documents perennially relevant? Are there common themes, methodologies, or breakthroughs that transcend temporal and disciplinary boundaries?

Finally, our exploration of the Collaboration Network, visualized in Figure 19, brought to light the interconnectedness of authors within the Trustworthy AI research landscape. The network metrics, including betweenness, closeness, and PageRank, shed light on influential nodes and clusters. Exploring the dynamics of collaboration networks prompts contemplation on the nature of interdisciplinary collaborations. How do these networks contribute to the synthesis of diverse perspectives, methodologies, and expertise within Trustworthy AI research?

In conclusion, our comprehensive analysis of Trustworthy AI research reveals a dynamic and interconnected ecosystem. The synthesis of findings from different sections offers a multifaceted understanding of the field. As we look toward the future, several avenues for research become apparent. Exploring the causative factors behind the observed patterns, investigating the qualitative aspects of highly cited works, and delving into the socio-economic determinants shaping global scientific production present exciting opportunities for future scholars.

In the grand tapestry of Trustworthy AI research, our analysis serves as a roadmap for navigating the complex interplay of authorship, impact, affiliations, global distribution, and collaborative networks. As we take stock of our journey, we leave the reader with a poignant reflection: Trustworthy AI is not merely a technological pursuit; it is a collective endeavor shaped by diverse minds, institutions, and nations. Embracing this diversity is key to unraveling the full potential of Trustworthy AI, ensuring that its benefits are realized globally and sustainably.

5. CONCLUSION

Our extensive investigation, into Trustworthy AI research conducted with a methodology based on principles has yielded valuable insights that have significant implications for the field. By using Lotkas Law and Bradfords Law we discovered patterns in author productivity and the distribution of work. This shed light on the spread of research output and the

concentration of sources. Examining factors such as authors local impact, affiliations and corresponding authors countries allowed us to identify influencers and understand how institutional and geopolitical factors shape the landscape of Trustworthy AI research. Additionally our analysis of production and citation impact provided insights into variations and the interplay between scientific output and influence. By visualizing the collaboration network among researchers our findings emphasized the importance of efforts in advancing this field. Collectively these outcomes contribute to a nuanced understanding of Trustworthy AI research establishing a foundation for exploration. Moving forward it is crucial to delve into aspects of highly cited works unravel socio economic determinants that influence global scientific production and explore dynamics within interdisciplinary collaborations in the Trustworthy AI landscape. The synthesis of our results highlights the necessity for an inclusive approach to Trustworthy AI research that recognizes the contributions, from various individuals, institutions and nations shaping its trajectory.

6. DATA SOURCE

The data, for this analysis on Trustworthy AI was gathered from the BibTeX file found in a GitHub repository. The dataset consists of factors, such as author details, document information, country specific scientific output, citation metrics and collaboration networks. This diverse and comprehensive dataset forms a basis for exploring and understanding the landscape of Trustworthy AI research. It offers insights into patterns of authorship, global contributions and significant affiliations. The meticulous curation of this dataset ensures its reliability and accuracy enabling an examination of the evolving trends and dynamics, within the field of Artificial Intelligence.

Funding

This study was carried out without receiving financial support from any institutions or sponsors.

Conflicts of Interest

The authors declare no conflicts of interest associated with this research.

Acknowledgment

The authors express heartfelt thanks to the institution for providing steadfast moral support and encouragement throughout the course of this research.

References

[1] D. Kaur, S. Uslu, K. J. Rittichier, and A. Durresi, "Trustworthy artificial intelligence: a review," *ACM computing surveys (CSUR),* vol. 55, no. 2, pp. 1-38, 2022.

[2] B. Li *et al.*, "Trustworthy AI: From principles to practices," *ACM Computing Surveys,* vol. 55, no. 9, pp. 1-46, 2023.

[3] R. Chatila *et al.*, "Trustworthy ai," *Reflections on artificial intelligence for humanity,* pp. 13-39, 2021.

[4] H. Liu *et al.*, "Trustworthy ai: A computational perspective," *ACM Transactions on Intelligent Systems and Technology,* vol. 14, no. 1, pp. 1-59, 2022.

[5] M. G. Yaseen and A. Albahri, "Mapping the evolution of intrusion detection in big data: a bibliometric analysis," *Mesopotamian Journal of Big Data,* vol. 2023, pp. 138-148, 2023.

[6] J. S. de Cerqueira, K.-K. Kemell, R. Rousi, N. Xi, J. Hamari, and P. Abrahamsson, "Mapping Trustworthiness in Large Language Models: A Bibliometric Analysis Bridging Theory to Practice," *arXiv preprint arXiv:2503.04785,* 2025.

[7] A. Kale, T. Nguyen, F. C. Harris Jr, C. Li, J. Zhang, and X. Ma, "Provenance documentation to enable explainable and trustworthy AI: A literature review," *Data Intelligence,* vol. 5, no. 1, pp. 139-162, 2023.

[8] R. Charter, "Ethical and trustworthy Artificial Intelligence," Tech. Rep2022.

[9] O. Freiman, "Making sense of the conceptual nonsense 'trustworthy AI'," *AI and Ethics,* vol. 3, no. 4, pp. 1351-1360, 2023.

[10] S. T. Mueller, R. R. Hoffman, W. Clancey, A. Emrey, and G. Klein, "Explanation in human-AI systems: A literature meta-review, synopsis of key ideas and publications, and bibliography for explainable AI," *arXiv preprint arXiv:1902.01876,* 2019.

[11] C. Lahusen, M. Maggetti, and M. Slavkovik, "Trust, trustworthiness and AI governance," *Scientific Reports,* vol. 14, no. 1, p. 20752, 2024.

[12] N. Tamascelli, A. Campari, T. Parhizkar, and N. Paltrinieri, "Artificial Intelligence for safety and reliability: A descriptive, bibliometric and interpretative review on machine learning," *Journal of Loss Prevention in the Process Industries,* p. 105343, 2024.

[13] N. Smuha, "Trustworthy artificial intelligence in education: Pitfalls and pathways," 2020.

[14] S. MARRONE, "TRUSTWORTHY AI: THE DEEP LEARNING PERSPECTIVE."

[15] A. K. Kaleel, A. Assad, and M. Fyadh, "Quranic Studies Evolution: A Bibliometric Analysis from 1880 to 2023," *Mesopotamian Journal of Quran Studies,* vol. 2024, pp. 1-12, 2024.

[16] G. Rieder, J. Simon, and P.-H. Wong, "Mapping the stony road toward trustworthy AI: Expectations, problems, conundrums," *Machines We Trust: Perspectives on Dependable AI. Cambridge, MA: MIT Press, Forthcoming,* 2020.
[17] A. Benseng and S. Tam, "Mapping the Evolution of Arabic Language Research: A Bibliometric Approach," *Mesopotamian Journal of Arabic Language Studies,* no. 2024, pp. 1-15, 2024.
[18] O. Albahri and A. AlAmoodi, "Navigating the metaverse of big data: A bibliometric journey," *Mesopotamian Journal of Big Data,* vol. 2023, pp. 92-106, 2023.
[19] M. Aljanabi, "Navigating the landscape: A comprehensive bibliometric analysis of decision-making research in civil engineering," *Mesopotamian Journal of Civil Engineering,* vol. 2023, pp. 35-45, 2023.
[20] M. Benk, S. Kerstan, F. von Wangenheim, and A. Ferrario, "Twenty-four years of empirical research on trust in AI: a bibliometric review of trends, overlooked issues, and future directions," *AI & society,* pp. 1-24, 2024.
[21] M. G. Yaseen and S. S. Alnakeeb, "Exploring the evolution of ai integration in English as a foreign language education: A Scopus-based bibliometric analysis (1997-2023)," *Mesopotamian Journal of Computer Science,* vol. 2023, pp. 143-158, 2023.