algorithms

MDPI

Article

Numbers Do Not Lie: A Bibliometric Examination of Machine Learning Techniques in Fake News Research

Andra Sandu [1], Ioana Ioanăș [2], Camelia Delcea [1], Margareta-Stela Florescu [3] and Liviu-Adrian Cotfas [1,*]

[1] Department of Economic Informatics and Cybernetics, Bucharest University of Economic Studies, 010552 Bucharest, Romania
[2] Department of Business and Administration, University of Bucharest, 030018 Bucharest, Romania
[3] Department of Administration and Public Management, Bucharest University of Economic Studies, 010374 Bucharest, Romania
* Correspondence: liviu.cotfas@ase.ro; Tel.: +40-771269599

Abstract: Fake news is an explosive subject, being undoubtedly among the most controversial and difficult challenges facing society in the present-day environment of technology and information, which greatly affects the individuals who are vulnerable and easily influenced, shaping their decisions, actions, and even beliefs. In the course of discussing the gravity and dissemination of the fake news phenomenon, this article aims to clarify the distinctions between fake news, misinformation, and disinformation, along with conducting a thorough analysis of the most widely read academic papers that have tackled the topic of fake news research using various machine learning techniques. Utilizing specific keywords for dataset extraction from Clarivate Analytics' Web of Science Core Collection, the bibliometric analysis spans six years, offering valuable insights aimed at identifying key trends, methodologies, and notable strategies within this multidisciplinary field. The analysis encompasses the examination of prolific authors, prominent journals, collaborative efforts, prior publications, covered subjects, keywords, bigrams, trigrams, theme maps, co-occurrence networks, and various other relevant topics. One noteworthy aspect related to the extracted dataset is the remarkable growth rate observed in association with the analyzed subject, indicating an impressive increase of 179.31%. The growth rate value, coupled with the relatively short timeframe, further emphasizes the research community's keen interest in this subject. In light of these findings, the paper draws attention to key contributions and gaps in the existing literature, providing researchers and decision-makers innovative viewpoints and perspectives on the ongoing battle against the spread of fake news in the age of information.

Keywords: fake news; social media; bibliometric analysis; n-gram analysis; machine learning; artificial intelligence; bibliometrix

Citation: Sandu, A.; Ioanăș, I.; Delcea, C.; Florescu, M.-S.; Cotfas, L.-A. Numbers Do Not Lie: A Bibliometric Examination of Machine Learning Techniques in Fake News Research. *Algorithms* **2024**, *17*, 70. https://doi.org/10.3390/a17020070

Academic Editors: Frank Werner and Fabrizio Marozzo

Received: 30 December 2023
Revised: 24 January 2024
Accepted: 2 February 2024
Published: 5 February 2024

1. Introduction

The explosive development of technology and the rising popularity of social media platforms have created ideal circumstances for the quick spread of fake news, which is nothing but altered data with the intention of confusing readers and influencing public opinion [1–3].

As a starting point for the discussion, one must know the differences between three popular terms that are somehow correlated, but they have different meanings—fake news, misinformation, and disinformation.

The invention and spread of false information with the aim to mislead, frequently in order to grab attention or manipulate public opinion, is known as fake news [4]. On the other side, misinformation is false information that is disseminated accidentally or through misunderstandings [5,6]. A more nuanced form is called disinformation, which is the intentional spread of incorrect information with the aim of misleading and influencing public opinion, frequently via well-planned and calculated initiatives [5].

Nevertheless, as Carmi et al. [4] pointed out, while the initial intention behind the emergence of the term "fake news" was to encompass in one concept both the misinformation and the disinformation terms, it has been observed that in contemporary discourse, certain political actors have appropriated the term as a means of discrediting news sources misaligned with their political perspectives. Consequently, the usage of the term "fake news" has led to significant confusion [5]. Due to this ambiguity, the Taylor & Francis website [7] offers a more in-depth discussion of these terms, while Lazer et al. [8] provide insightful discourse on the science of fake news.

Thus, it has been noted that the key differentiators which highlight the significance of media literacy and critical thinking in navigating the complicated information world and differentiating fact from fiction are considered to be intention, awareness, and purpose [9–11].

Fake news impacts individuals in numerous manners and has an enormous effect on modern society. This false data has the power to manipulate people's opinions and beliefs, which may consequently contribute to the development of wrong attitudes and conflicts within society. They may amplify emotions such as fear, anxiety, and fury, which can have negative implications for people's mental health. Additionally, fake information has the potential to damage public confidence in democratic institutions and the media, harming a society that depends on accurate information for its proper functioning as an entire system.

Machine learning plays a crucial role in reducing the spread of fake news by using complex algorithms to examine huge volumes of textual data [12]. These algorithms, which have been trained on a variety of datasets, recognize language nuances and patterns that suggest false information, and their accuracy is improved by natural language processing techniques, which enable dynamic adaptability to changing disinformation strategies [13]. By empowering platforms and consumers to make informed and educated choices regarding the reliability of news sources, this technology helps to strengthen the fight against the spread of false information.

Those subjects have attracted the interest of many researchers that have conducted multiple investigations, such as, but not limited to, combating fake news with transformers [14], providing an automated classification of fake news spreaders for the purpose of breaking the misinformation chain [15], detecting fake news through the use of machine learning and deep learning [16], employing automatic fake news detection in the case of online news [17], using a feature-centric classification approach that integrates both ensemble learning and deep learning methods for fake news detection [13], and predicting the evolution of news spread [18] or broader subjects related to discussing the trends and challenges in identifying fake news on social networks based on natural language processing [19].

The COVID-19 outbreak represents just one recent example of an event that generated a wave of fake news [20]. Not only has the COVID-19 pandemic caused tensions to rise, but, additionally, there has also been a shocking explosion in false information and fake news, which has had considerable consequences on how the public reacts to the virus. Conspiracy theories questioning the virus's origin and the efficacy of vaccinations, as well as risky advice and unconfirmed scientific treatments for the illness, have all been included in COVID-19 fake news [21]. The misunderstanding brought on by this fake news has made it difficult for people to communicate clearly and take preventative action, and among the negative effects are poor confidence in trustworthy sources of information, unwillingness to be vaccinated, disapproval of public health initiatives, and, in certain situations, unjustified fear. This subject was debated and analyzed by multiple researchers, who have tried to address the fake news phenomenon in the case of COVID-19 pandemic from multiple perspectives, such as, but not limited to, using the pre-bunking (psychological inoculation theory) as an efficient solution for increasing the resistance in large populations to fake news [22], proposing a fake news detector in the case of COVID-19 by mixing named entity recognition and stance classification [23], discussing the predictors of fake news sharing

in the case of social media users [24], and identifying the trends in fake news in times of COVID-19 [25].

Given that fake news phenomenon poses a tangible threat in today's interconnected society, education and the cultivation of professional journalism have emerged as pivotal tools in the ongoing fight against it. Understanding this phenomenon, identifying its origins, and implementing effective countermeasures are crucial steps to safeguard truth, democracy, and information integrity in our contemporary era.

As can be deduced from the intriguing title of the article, the purpose of the work conducted in this paper is to bring to the fore a bibliometric perspective in the cutting-edge area of machine learning techniques in fake news research, including comprehensive analysis in terms of authors, sources, words, countries, universities, and many more, in order to discover trends, insights, themes, and opportunities, and develop strategies for combating the spread of fake news in our age of information. This study sheds light on the key contributions existing in the current scientific literature, offering crucial details in this area by analyzing numerical values, important indicators, tables, graphs, and visual representations based on the extracted papers.

The contribution of this work is truly important for the scientific community, since it offers an objective point of view of the current literature, regarding popular sources and highly cited articles, the leading contributors, and the most prolific authors, affiliations, etc. Apart from this, it also delves into interesting subjects such as deep word analysis, collaboration maps, networks, and the list goes on with many more topics, highlighted in Section 3.

Having said this, the aim of the article is to answer to a series of questions such as the ones listed below:

- Q1: what is the main information about the extracted articles used in this investigation and what are the values for the most relevant bibliometric indicators?
- Q2: who are the most prolific authors who have published papers in the area of machine learning techniques in fake news research?
- Q3: which are the popular sources and most cited articles, what are the methods used, and what is the purpose?
- Q4: which are the affiliations that registered the highest number of citations in the area of machine learning techniques in fake news research?
- Q5: which are the countries marked as leading contributors, and what insights can be observed in terms of collaborations?
- Q6: what are the findings related to collaboration networks in this domain?
- Q7: what are the conclusions drawn from the word analysis?

In order to provide answers to all of these questions, and not only those, we extracted a dataset collection of papers in the area of machine learning techniques in fake news research from the WoS database.

The current article has a clear structure: Section 2 consists of data about materials and methods, Section 3 includes the dataset analysis through multiple bibliometric indicators, Section 4 presents the discussions, Section 5 outlines the limitations, and, last but not least, Section 6 summaries all the findings and draws attention to the most important insights.

2. Materials and Methods

The present analysis benefits from the comprehensive access to a vast array of academic literature facilitated by the utilization of the Clarivate Analytics' Web of Science Core Collection, formerly known and referred to as Web of Science (WoS) [26]. This platform served as the primary tool for curating a corpus of papers essential for conducting a rigorous bibliometric examination of machine learning techniques in the context of fake news research.

Facilitating collaboration and fostering innovation, the WoS database stands as an indispensable instrument in advancing academic endeavors and enriching the collective knowledge base within scholarly circles. Serving as a pivotal gateway, it functions as a

central hub, expediting access to a broad spectrum of scientific publications. The platform streamlines researcher browsing, retrieval, and a dissemination of findings, enhancing efficiency through its expansive coverage across diverse domains. Offering reliable and comprehensive means to stay abreast of current developments, engage in interdisciplinary research, and make substantial contributions to the body of knowledge, this platform emerges as a critical resource for academics.

Despite the existence of other popular databases, such as Scopus or IEEE, WoS seems to be holding a prominent position in the scientific community. WoS works on a subscription basis, offering users personalized access. The importance of full access to sources used in bibliometric analysis (as is the case of this article) is addressed in the paper of Liu [27]. Furthermore, as Bakir et al. [28] pointed out, the WoS platform provides a higher level of coverage when considering the variety of the disciplines included in this database, while being considered at the same time the most credible by the scientific community, even though it is less inclusive than its counterparts. These aspects are further highlighted in the works of Cobo et al. [29] and Modak et al. [30]. Nevertheless, with the possibility of directly importing raw data extracted from WoS into Biblioshiny, the R tool used in the present analysis was a key factor in our choice [31].

Given that the WoS database operates on a subscription-based model, scholars such as Liu [32] and Liu [27] underscore the criticality of explicitly delineating, at the outset of a bibliometric analysis, the specific indices to which the individual conducting the database extraction has been granted access. Notably, it has been observed that variations in the obtained dataset can arise due to the level of subscription, thereby emphasizing the need for transparency in disclosing the scope of database access. In the present study, it is imperative to affirm that comprehensive access was secured to all ten indexes provided by the WoS platform.

Due to all the above-mentioned features and doubled by the use of the WoS platform in similar bibliometric analyses on different themes and areas, we decided to opt for this database when extracting our dataset [28–30,33].

Additionally, we based our decision in using solely the WoS database (and not a combination of two or more databases) on two other aspects.

The first one was represented by a search of the bibliometric papers that have used WoS as a primarily database, and it has been observed that from the 28,708 papers featuring "bibliometric" as a keyword, a higher number of papers mention the WoS database (namely, 9292 papers) when compared to the number of papers that mention Scopus (namely, 5573 papers)—please consider Table 1 for the query we have used (the search was performed on 24 January 2024). Moreover, if we exclude the "scopus" keyword from the title/abstract/authors' keywords of the 9292 papers that mention WoS in the title/abstract/authors' keywords, it can be observed that 7622 papers refer to solely the WoS database. In either case, WoS seems to be the most prominent platform when conducting bibliometric analysis, and mixing the information from two different databases seems to be a limited practice among the researchers conducting bibliometric analyses.

The second aspect for conducting the analysis on a single database is related to the type of individual analysis included in the bibliometric analysis—for example, as we provide a list of the most highly cited papers, it is not clear how we should have conducted this analysis if a paper was in both databases with respect to the number of citations. It is known that WoS and Scopus offer the number of citations per paper based on their own records, providing only a part of the story as there are journals indexed in both databases, but there are also journals indexed in one or in the other database. An alternative would have been to sum them up, but this situation would have favorized the papers published in journals indexed in both databases.

Thus, considering the researchers mentioned above that have advocated for the use of the WoS database in bibliometric studies, together with our own search in terms of the number of papers in each category, we have decided to use only the WoS database for performing the analysis included in the paper.

Table 1. Exploration on selecting the database.

Exploration Steps	Filters on Web of Science	Description	Query	Query Number	Count
1	Keywords	Contains specific keywords related to bibliometric analysis in title/abstract/authors keywords	((TI = ("bibliometric")) OR AB = ("bibliometric")) OR AK = ("bibliometric")	#1	28,708
		Contains specific keywords related to WoS database in title/abstract/authors keywords	(((((TI = (WoS)) OR AB = (WoS)) OR AK = (WoS)) OR TI = (Web_of_science)) OR AB = (Web_of_science)) OR AK = (Web_of_science)	#2	100,399
		Contains specific keywords related to Scopus database in title/abstract/authors keywords	((TI = ("Scopus")) OR AB = ("Scopus")) OR AK = ("Scopus")	#3	75,325
2	Title/Abstract/ Author's Keywords	Contains one of the bibliometric analyses and WoS-specific keywords	#1 AND #2	#4	9292
		Contains one of the bibliometric analyses and Scopus-specific keywords	#1 AND #3	#5	5573
		Contains one of the bibliometric analyses and WoS-specific keywords but does not contains Scopus keywords	(#4) NOT TI = (Scopus) NOT AB = (Scopus) NOT AK = (Scopus)	#6	7622

However, we are aware of the limitations of relying solely on one database, including aspects such as incomplete coverage and publication lag, and we strongly agree that the use of multiple sources for paper extraction might have been conducted for a slightly different dataset.

As suggested by Marin-Rodriguez et al. [34], the analysis is divided into two main stages: the dataset extraction and the bibliometric analysis, as depicted in the following.

2.1. Dataset Extraction

The dataset extraction process was carefully explained below, divided into five steps, as can be observed in Table 2.

The first exploratory step was comprised of three different queries. The first query had the purpose of identifying the finer points of false information with a concentrated search approach that examined titles, abstracts, and author keywords for the term "fake news". This produced a corpus of 5671 documents, which established the foundation for comprehending the academic conversation around the fake news subject.

The research delved into the broad subject of deep learning and machine learning, making this an ideal shift. The terms "machine_learning" and "deep_learning" in titles, abstracts, or author keywords were the focus of two simultaneous searches that were conducted. A remarkable number of 352,001 papers for machine learning and 223,762 documents for deep learning, were produced as a result of this dual strategy, serving as proof that the contributions of artificial intelligence (AI)-related research present a significant influence on the academic discussion.

The outline of the process took on a deeper significance, since it explored the nexus between deep learning and machine learning, merging papers from the two domains to create a single dataset with 532,179 papers. This combination shed light on the mutually beneficial link between academic research and technology advancements and allowed for a more in-depth investigation of the convergence of technological capabilities and scholarly discourse.

Table 2. Data selection steps.

Exploration Steps	Filters on Web of Science	Description	Query	Query Number	Count
1	Keywords	Contains specific keywords related to fake news in title/abstract/authors keywords	((TI = ("fake_news")) OR AB = ("fake_news")) OR AK = ("fake_news")	#1	5671
		Contains specific keywords related to machine learning in title/abstract/authors keywords	((TI = ("machine_learning")) OR AB = ("machine_learning")) OR AK = ("machine_learning")	#2	352,001
		Contains specific keywords related to machine learning in title/abstract/authors keywords	((TI = ("deep_learning")) OR AB = ("deep_learning")) OR AK = ("deep_learning")	#3	223,762
2	Title/Abstract/ Author's Keywords	Contains one of the machine learning or deep learning specific keywords	#2 OR #3	#4	532,179
		Contains one of the machine learning- or deep learning-specific keywords and fake news keywords	#1 AND #4	#5	900
3	Language	Limit to English	(#5) AND LA = (English)	#6	897
4	Document Type	Limit to Article	(#6) AND DT = (Article)	#7	510
5	Year published	Exclude 2023	(#7) NOT PY = (2023)	#8	347
		Exclude 2024	(#8) NOT PY = (2024)	#9	346

The second exploratory step had the purpose of further clarifying the narrative. In order to achieve this, the papers that covered both the fake news area and the intersection of machine learning and deep learning were identified afterwards, resulting in a collection of 900 distinct publications, offering an innovative viewpoint on the complex interactions between these distinct fields.

The third exploratory step followed a precision-oriented approach to the linguistic part of the investigation, restricting the attention to papers written in the English language—this choice was in line with the fact that English has been considered an exclusion criteria in other bibliometric works, such as the ones conducted by Stefanis et al. [35], Gorski et al. [36], and Fatma and Haleem [37]. After linguistic curation, the ensemble was reduced to 897 articles, which ensured consistency and clarity in the next stages of the current study. This small difference of only 3 articles being excluded from the analysis after applying the language restriction suggests the fact that the vast majority of researchers opt for writing articles in this area in English, being the most well-known language among the scientific community.

The document types occupying the center of attention in the fourth exploratory step followed in the selection of the data set. Therefore, the analysis focused on the exclusive selection of works marked as articles, a restriction that led to a significant decrease in the number of documents, namely 510. Here, it shall be noted that the inclusion of a paper into the "article" category by WoS is based on the fact that, according to the WoS, the paper provides new and original work [38]. As a result, in this category, WoS includes research papers, brief communications, technical notes, chronologies, full papers, and case reports that were published in a journal and/or presented at a symposium or conference, as well as proceedings papers [38]. We have added this note as it is well stated in the scientific literature the importance of the document type when conducting bibliometric analyses [39]. This exclusion criterion has been used also by Fatma and Haleem [37] in their research.

With the goal to conduct the final exploratory stage of the selection process, the years 2023 and 2024 were excluded as a time limitation. By restricting the number of articles evaluated in this study to those published over a 6-year time limit, from 2017 to 2022, this temporal cut guaranteed that our analysis was grounded in a precise time span. As a result, the corpus was subsequently reduced to 346 articles.

2.2. Bibliometric Analysis

In order to conduct the bibliometric analysis, the chosen data set was thoroughly examined using Biblioshiny [40], a well-known research tool included in R studio. With its assistance, a sizable number of visually appealing and functional graphs, tables, and visualizations were retrieved, emphasizing crucial information about the investigation of machine learning approaches in fake news research as well as hidden features, unknown details, current trends, and tactics [41,42].

The bibliometric analysis was conducted by following five distinctive facets, namely the information regarding the overview of the dataset, the analysis of the sources, the analysis of the authors, the analysis of the papers included in the dataset, and a mixed analysis [43].

The initial facet was devoted to providing an overview of the dataset with the aim of presenting a broad perspective on its size and general attributes related to authors, papers, and sources. These aspects would subsequently be subjected to a more detailed analysis in the ensuing sections of the paper. The elements included and discussed in this facet are listed in Table 3.

Table 3. Elements included in dataset overview.

Dataset Overview	
Timespan	Author's keywords
Sources	Authors
Documents	Author appearances
Average years from publication	Authors of single-authored documents
Average citations per documents	Authors of multi-authored documents
Average citations per year per document	Single-authored documents
References	Documents per author
Annual scientific production evolution	Authors per document
Annual average article citations per year evolution	Co-authors per documents
Keywords plus	Collaboration index

Another facet was dedicated to the analysis of the journals in which the authors had decided to publish, highlighting elements related to the quantity of the papers and their impact. The analysis encompasses similar elements to those provided in Table 4.

Table 4. Elements included in sources analysis.

Sources Analysis
Most relevant journals
Bradford's law on source clustering
Journals' impact based on H-index
Journals' growth (cumulative) based on the number of papers

The author analysis represents another facet in which the focus is on the most prominent authors and their characteristics, e.g., production over time, affiliations, countries, collaboration map and collaboration network—please consider Table 5.

Table 5. Elements included in author analysis.

Author Analysis	
Top authors based on number of documents	Scientific production based on country
Top authors' productions over time	Top countries with the most citations
Most relevant affiliations	Country collaboration map
Most relevant corresponding author's country	Top authors' collaboration network

The paper analysis focuses on the most globally cited documents through both a review and an overview. Also, a word analysis is included for better shaping the field of fake news and its associated implications. The elements provided in Table 6 are discussed in this section.

Table 6. Elements included in paper analysis.

Paper Analysis	
Most global cited documents—overview	Most frequent bigrams in abstracts and titles
Most global cited documents—review	Most frequent trigrams in abstracts and titles
Most frequent words in keywords plus	Co-occurrence network for the terms in authors' keywords
Most frequent words in authors' keywords	Thematic map the terms in authors' keywords
Top words based on keywords plus and authors' keywords	

The mixed analysis completes the bibliometric approach by providing connections between the other discussed elements (e.g., authors, affiliations, journals) through the use of three-fields plots.

3. Dataset Analysis

The final dataset relevant for the examination of machine learning techniques in fake news research, collected after applying all the criteria in the previous section, was deeply examined from multiple perspectives in the following pages, including in an advanced analysis regarding sources, authors, literature, words, and more.

3.1. Dataset Overview

The main information about the data can be found below, in Table 7. The examined collection assesses 346 papers from 175 different sources, in a time period which covers the years 2017 to 2022, indicating the novelty in both the utilization of the "fake news" term and in the recent interest of the research community to this subject when coupled with ML techniques. Considering the small value of 1.79 found for average years in publications, this highlighted the fact that most of the papers included in the analysis were recent, with the study being relevant for current challenges.

Table 7. Main information about data.

Indicator	Value
Timespan	2017:2022
Sources	175
Documents	346
Average years from publication	1.79
Average citations per document	16.21
Average citations per year per document	5.048
References	10,991

The academic significance and the long-lasting impact of machine learning techniques to the ever-evolving subject of fake news is demonstrated by the remarkable average of

16.21 citations per document, the 5.048 average citations per year per document, and the substantial number of references, namely 10,991.

Figure 1 shows how scientific production evolved annually during the analyzed timestamp, namely the period between 2017 and 2022. The trajectory shows a substantial upward trajectory in academic production. Starting insignificantly in 2017, with just one article published, the number of articles expanded steadily over the next several years, reaching 3 in 2018, 19 in 2019, 47 in 2020, and an exponential increase to 106 in 2021. The year 2022 was the highest point of this rising trend in scientific productivity, with 170 papers published, presenting a huge annual growth rate of 179.31%. This considerable rise can be attributed to the fact that fake news has become more widespread in recent years, drawing the interest of several researchers who have investigated machine learning techniques for identifying and combating this alarming phenomenon.

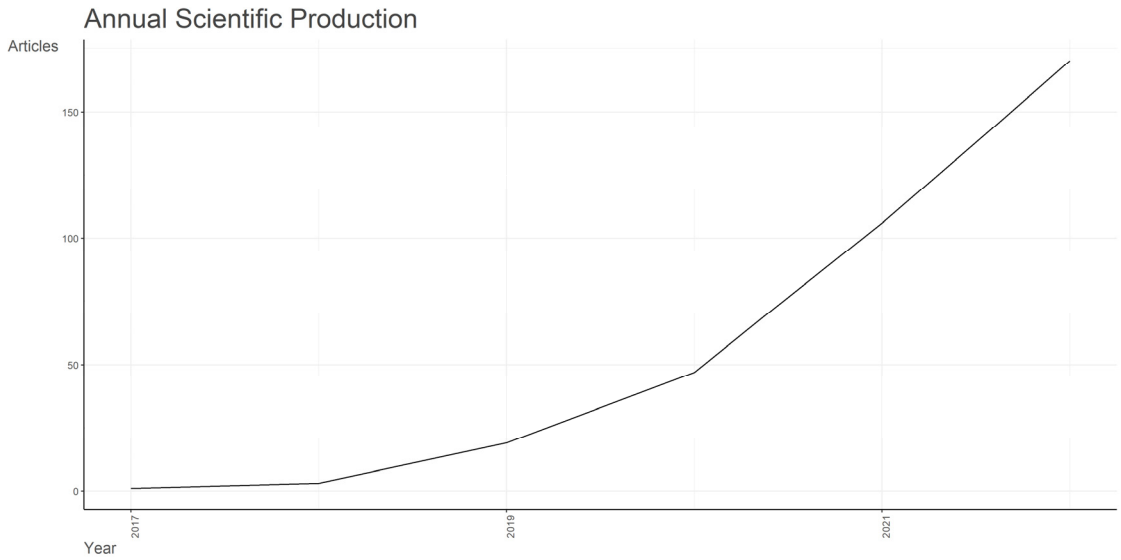


Figure 1. Annual scientific production evolution.

Figure 2 shows a changing pattern with oscillating values between 1.57 and 9.05, in terms of the annual average-article-citations-per-year evolution, examined between 2017 and 2022.

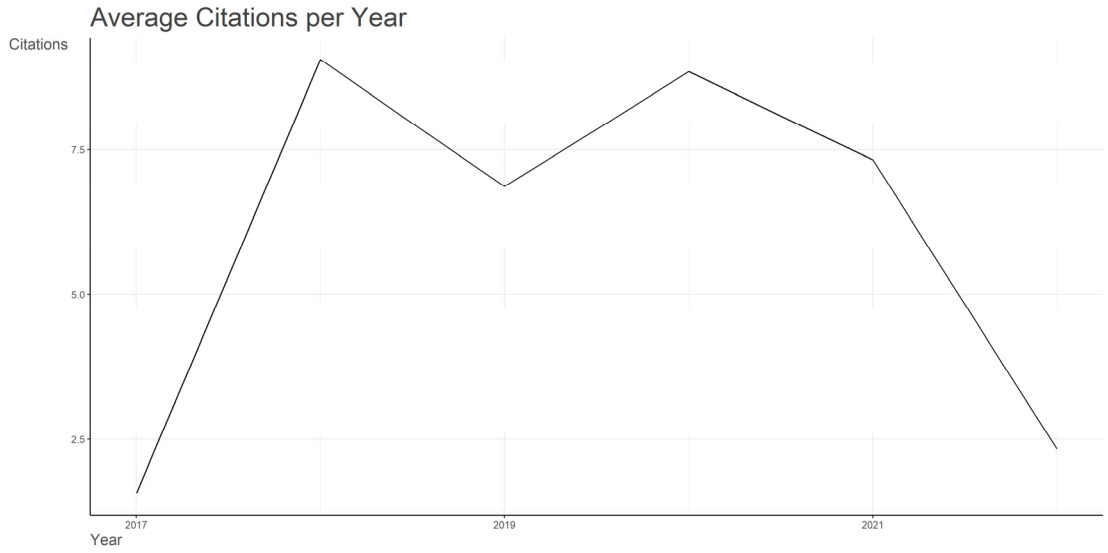


Figure 2. Annual average-article-citations-per-year evolution.

A low mean of 1.57 was reported in 2017, demonstrating early visibility, while a substantial rise of 9.05, the highest recorded value, occurred in 2018, proving increasing attention. The high averages recorded in the following years (2019, 2020, and 2021) remained in the top (6.86, 8.84, and 7.32), suggesting ongoing academic relevance. The mean did, however, drop to 2.33 in 2022, perhaps as a consequence of the limited amount of time between the moment the papers had been published and the moment in which the dataset had been extracted.

Essential information on the variety of words used is given through the indicators depicted in Table 8. With an average of 0.57 keywords per page, "Keywords Plus", also known as index terms, had quite a low value of 198, indicating the usage of a more focused vocabulary. The "Author's keywords" list has expanded to 861, suggesting a broader range of terms selected by the authors, resulting in a recorded average value of 2.49 of such terms per document.

Table 8. Document contents.

Indicator	Value
Keywords plus	198
Authors' keywords	861

The listed indicators presented in Table 9 shed light on the dynamics of authorship in the examined sample. A total of 1204 writers appeared 1329 times in the data set, meaning that writers were mentioned more than once across multiple publications. Furthermore, it is also noteworthy that 11 authors were identified as authors of single-authored articles, proving that single-authored academic publications do occur in the area of the examination of machine learning techniques in fake news research. However, the majority of authors, namely 1193, collaborated to write multi-authored publications.

Table 9. Authors.

Indicator	Value
Authors	1204
Author appearances	1329
Authors of single-authored documents	11
Authors of multi-authored documents	1193

Table 10 brings to the forefront details regarding author collaboration. The premise that few works are the result of individual contribution is supported by the small value registered for single-authored documents, namely 11. The value for documents per author index is 0.287, recorded because the number of extracted articles is higher than the number of authors (346 versus 1204). The steady trend in collaboration around the examination of machine learning techniques in fake news research is truly evident, with our hypothesis proven by the significant value of 3.48 registered for authors per document index, along with the dataset's 3.56 collaboration index, and the co-authors per document measure, which stands at 3.84, indicating that approximately four writers were involved in a single paper.

Table 10. Authors collaboration.

Indicator	Value
Single-authored documents	11
Documents per author	0.287
Authors per document	3.48
Co-authors per documents	3.84
Collaboration index	3.56

3.2. Sources

Figure 3 depicts the top 18 most relevant journals for the present analysis, and in order to achieve this, a pre-defined criterion was established whereby each journal needed to have at least four publications in the area of the examination of machine learning techniques in fake news research in order to be considered at the top.

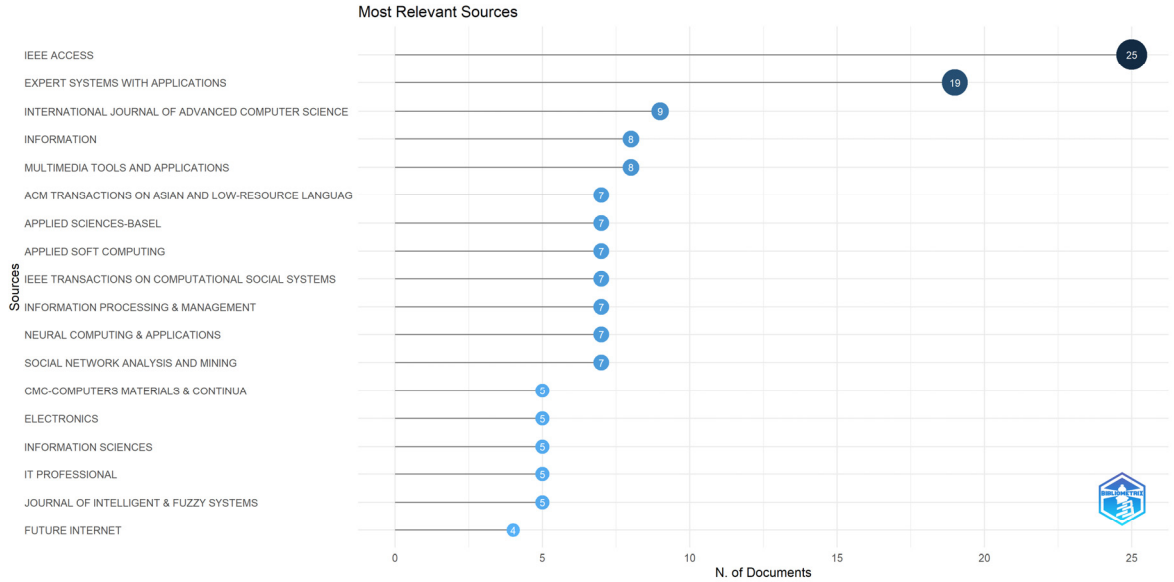


Figure 3. Top 18 most relevant journals.

As can be observed below, the leadership position is held by a popular journal, namely *IEEE Access*, with a significant number of 25 documents published in this area. The second place is occupied by another well-known journal—*Expert Systems with Applications*, with 19 documents, while *International Journal of Advanced Computer Science and Applications* is ranked in the third position with 9 documents.

Other relevant journals in this area of research are listed here in alphabetical order: *Information*, *Multimedia Tools and Applications*, *ACM Transactions on Asian and Low-Resource Language Information Processing*, *Applied Sciences-Basel*, *Applied Soft Computing*, *IEEE Transactions on Computational Social Systems*, *Information Processing & Management*, *Neural Computing & Applications*, *Social Network Analysis and Mining*, *CMC-Computers Materials & Continua*, *Electronics*, *Information Sciences*, *IT Professional*, *Journal of Intelligent & Fuzzy Systems*, and *Future Internet*.

Considering the variety of subjects they cover, including technology, engineering, computer science, multimedia, and artificial intelligence, these highly esteemed journals are thought to be suitable for publishing articles in the area of the examination of techniques for machine learning, thus making them useful for researchers investigating fake news.

Samuel C. Bradford came up with Bradford's Law, which clarifies the discrepancy in the distribution of scientific publications in a particular sector [44,45]. The aforementioned concept distinguishes the literature into three distinct components: core—Zone 1, with a few highly important journals, middle—Zone 2, relatively productive, and external—Zone 3, including a large number of less important journals. This paradigm makes it easier to identify relevant content and give special attention to certain subject areas [44,45].

Figure 4 draws the reader's attention to Bradford's law on source clustering, presenting the highest cited journals found in Zone 1, namely *IEEE Access, Expert Systems with Applications, International Journal of Advanced Computer Science and Applications, Information, Multimedia Tools and Applications, ACM Transactions on Asian and Low-Resource Language Information Processing*, *Applied Sciences-Basel*, *Applied Soft Computing*, *IEEE Transactions on Computational Social Systems*, *Information Processing & Management*, *Neural Computing & Applications*, and *Social Network Analysis and Mining*.

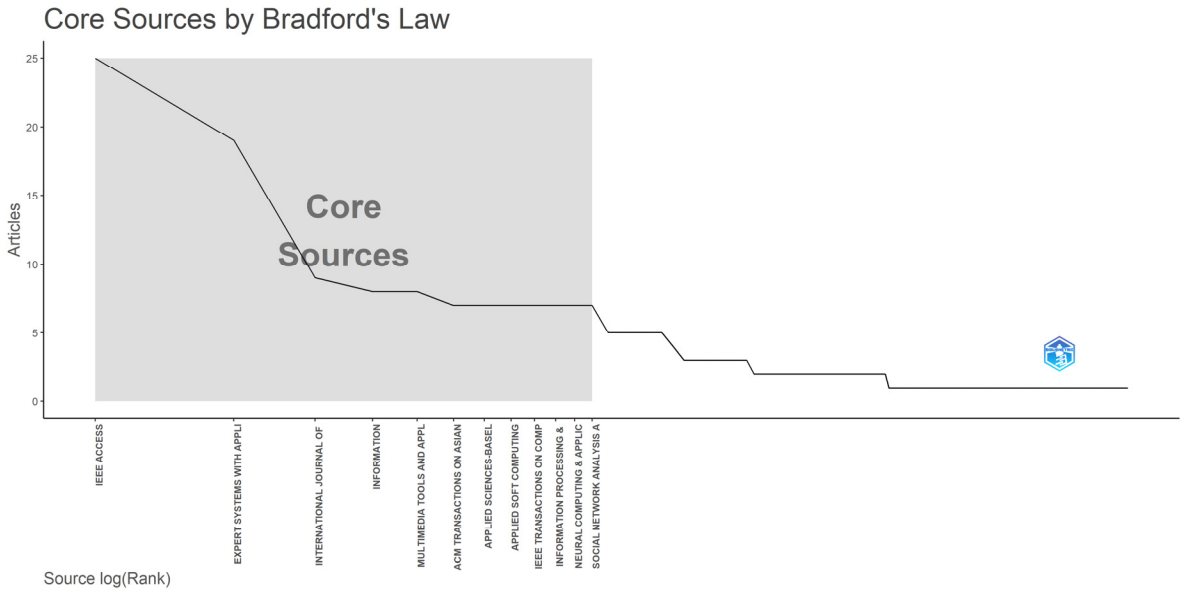


Figure 4. Bradford's law on source clustering.

Figure 5 presents the graphical view of the journal's impact based on the H-index. The H-index is a popular metric used for proving the significance of the papers, by measuring the total number of works that have been published by a specific journal and which have gathered at least some H-citations.

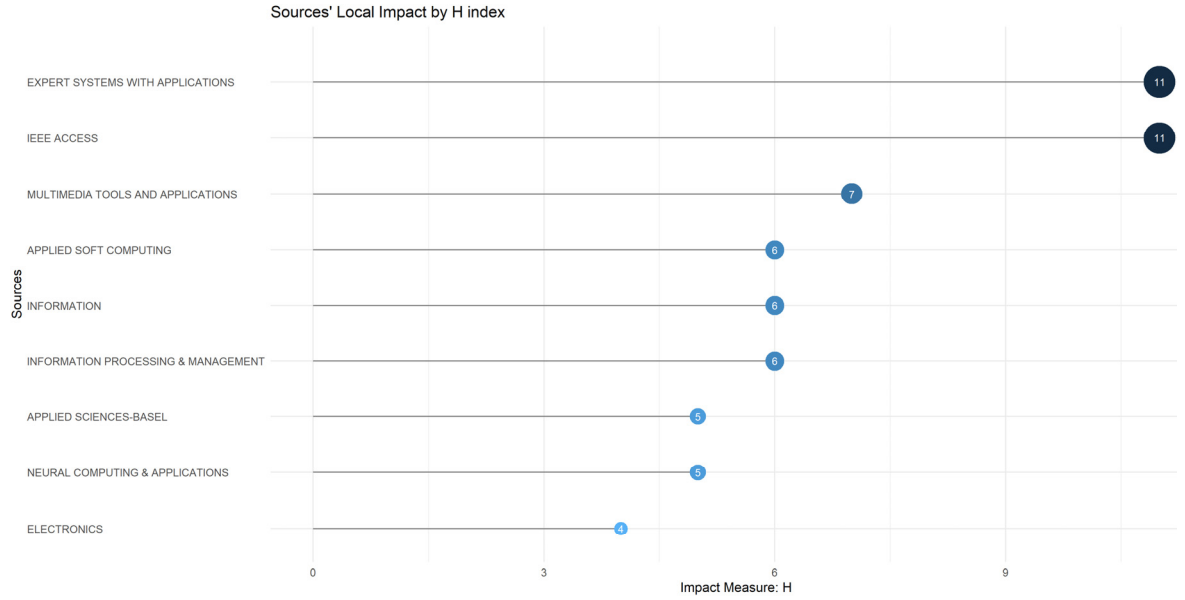


Figure 5. Journals' impact based on H-index.

The leadership position in top by considering the H-index indicator is held by two popular journals, listed alphabetically; namely, *Expert Systems with Applications* and *IEEE Access*, each with 11 papers and recording at least 11 citations in the area of the examination of machine learning techniques in fake news research. As expected, these journals also belong to Zone 1, according to Bradford's law—please see Figure 4.

Figure 6 sheds light on the sources' productions over time, considering the number of published papers. The journals *Expert Systems with Applications*, *IEEE Access*, *Information*, *International Journal of Advanced Computer Science and Applications*, and *Multimedia Tools and Applications*, as expected after the above analysis, reflect the most noteworthy growth.

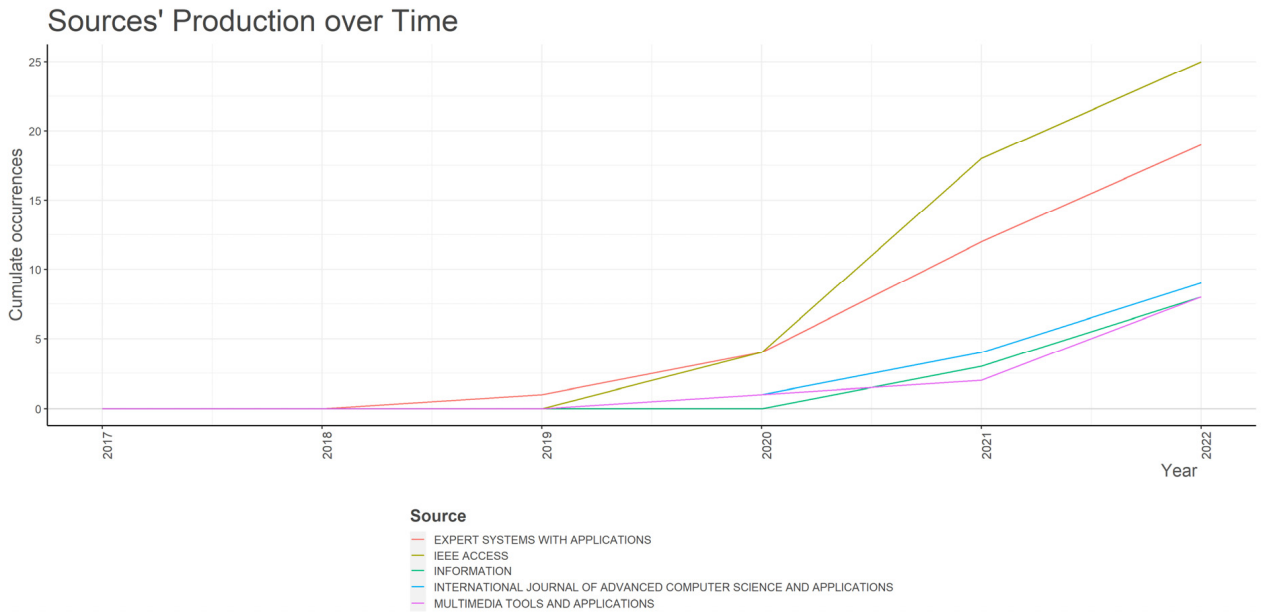


Figure 6. Journals' growth (cumulative) based on the number of papers.

3.3. Authors

Figure 7 presents the most influential authors accordingly the number of published papers related to the examination of machine learning techniques in fake news research. The top 21 writers were picked based on an established criterion, focusing on authors who had contributed to no fewer than three publications in the field being taken into consideration.

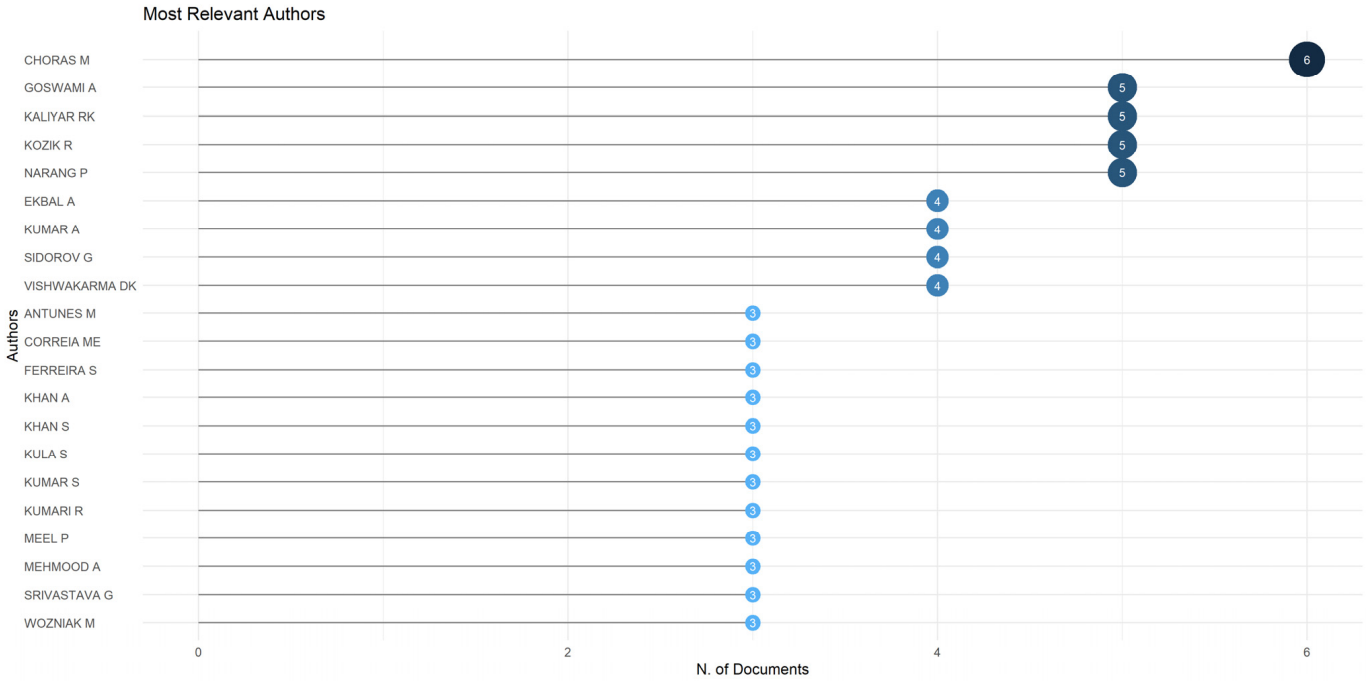


Figure 7. Top 21 authors based on number of documents.

The first place is occupied by Choras M, with six published articles, followed closely by other significant authors, alphabetically ordered as Goswami A, Kaliyar RK, Kozik R, and Narang P, each with five articles. For the entire list, please consult Figure 7.

In terms of authors' production over time, it is obvious to detect from Figure 8 a rise in interest and an increased number of papers published in the field of machine learning techniques in fake news research, starting with the year 2020.

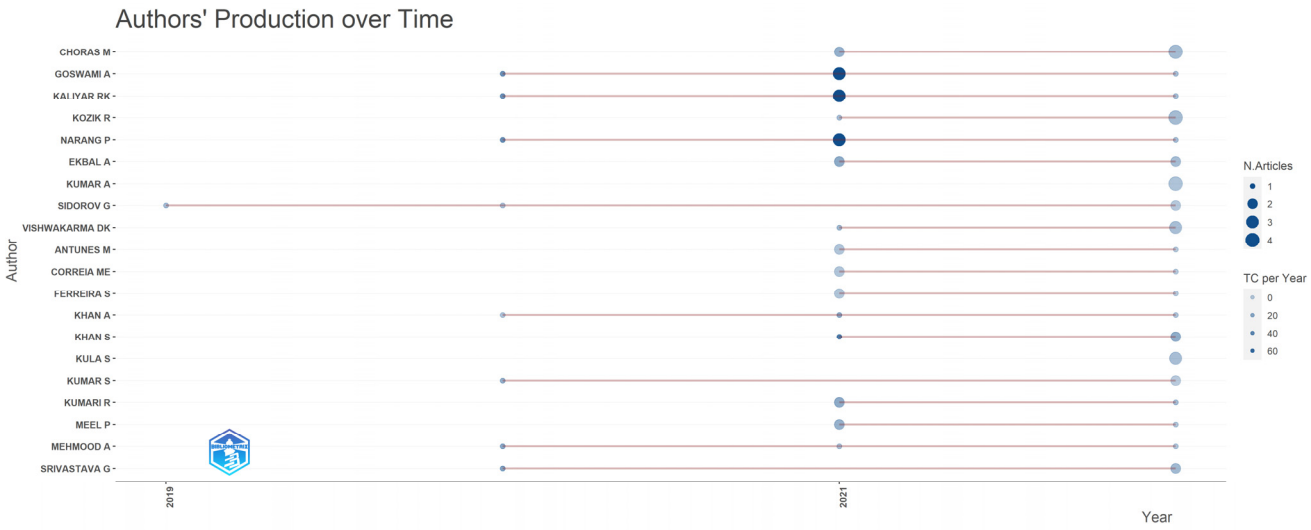


Figure 8. Top 20 authors' production over time.

This obvious boost could perhaps be associated with the worldwide outbreak of the COVID-19 virus that lead to huge increase in the amount of fake and alarming news, aimed to induce panic and fear among people across the globe. This worldwide event drew the attention of many researchers, who attempted to include in their research they were conducting various algorithms based on machine learning for identifying and preventing the spread of fake news over the internet, particularly on social media networks.

Figure 9 depicts the most relevant affiliations based on the number of published articles in the studied area— the examination of machine learning techniques in fake news research. The top affiliations were picked using a criterion that guaranteed inclusion for affiliations with at least five papers in this specific field.

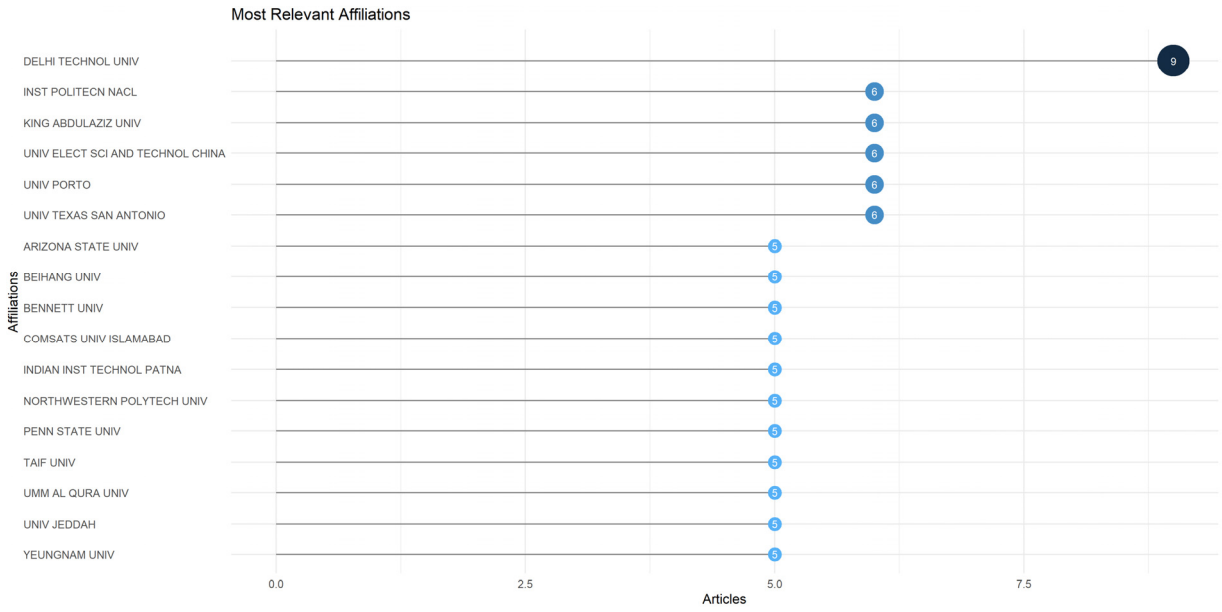


Figure 9. Top 17 most relevant affiliations.

The Delhi Technological University is ranked in first place, with a total of nine articles, followed closely by other famous affiliations—for the entire list, please see Figure 9.

Figure 10 brings to the foreground the top 19 most relevant corresponding authors' countries.

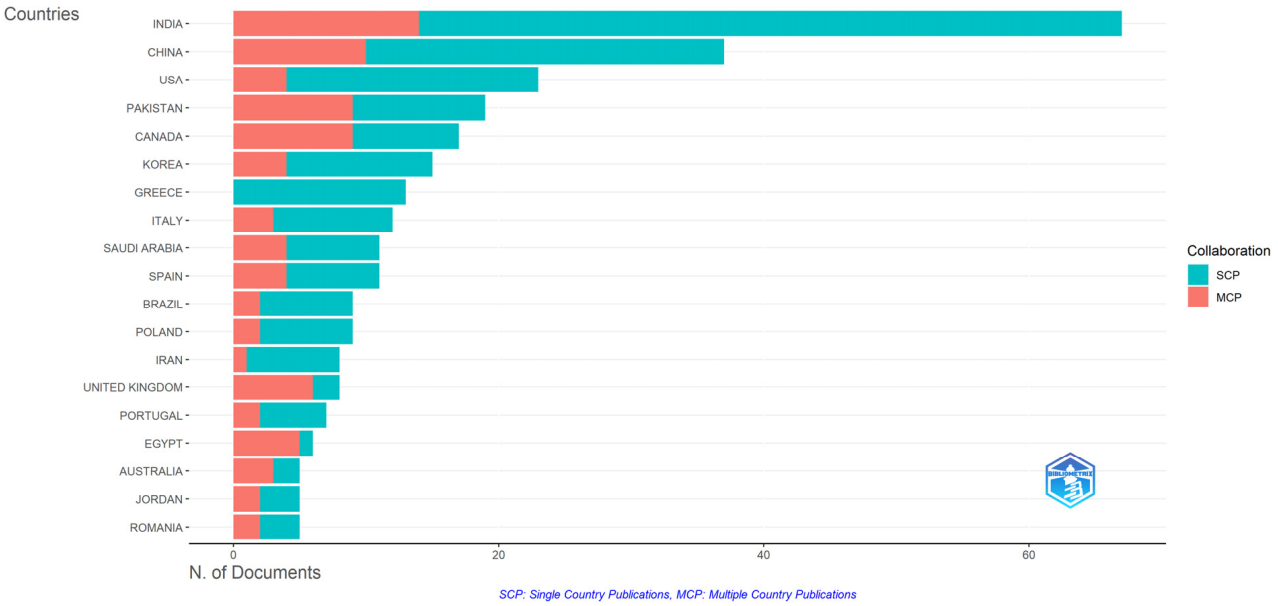


Figure 10. Top 19 most relevant corresponding authors' countries.

As can be clearly observed, the leadership position is held by India, with an impressive number of published articles—67, along with the highest scores registered for Single Country Publications and Multiple Country Publications (SCP = 53, MCP = 14). For more information, kindly refer to Figure 10.

Figure 11 illustrates a graphical depiction of the map that has been colored with respect to scientific productivity in the field of machine learning techniques in fake news research. A strong shade of blue indicates a substantial number of publications (India, China, and the United States), while a light gray color suggests an insignificant quantity of publications (Thailand, Singapore, and Norway).

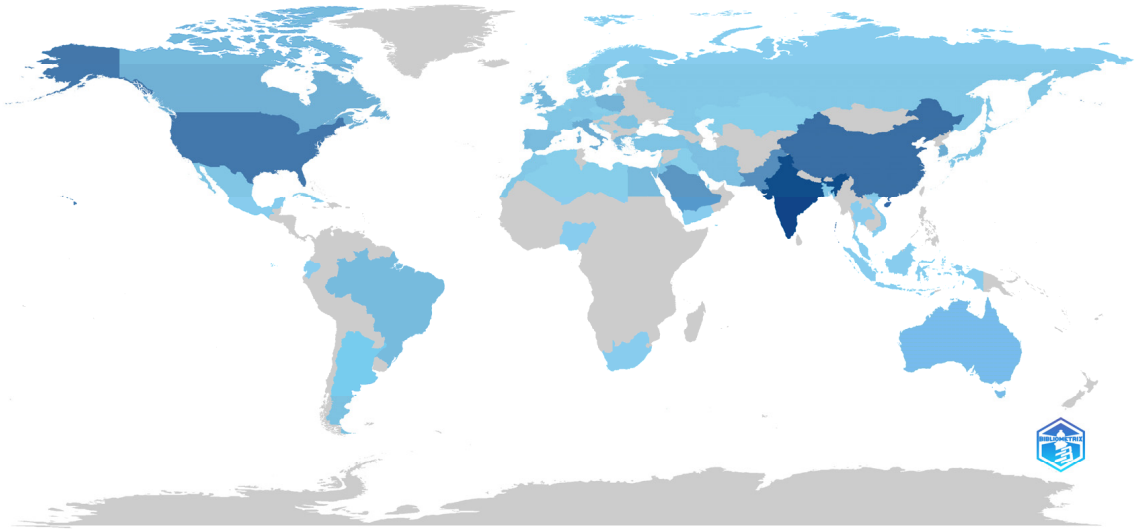


Figure 11. Scientific production based on country.

Figure 12 presents the top 20 countries with the most citations, and, as anticipated, the leader of the ranks is India, with a huge number of citations, namely 1249, along with a high value for average article citations: 18.60. In significant difference to first place, there are Canada and China. For the whole list, please see Figure 12.

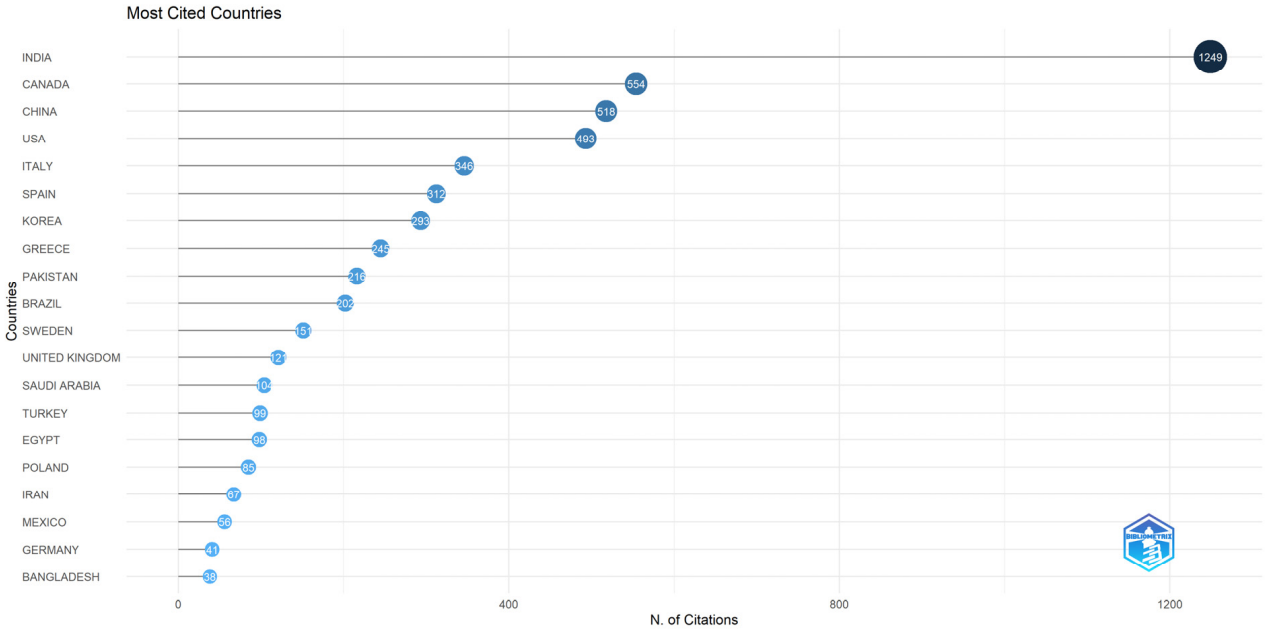


Figure 12. Top 20 countries with the most citations.

Figure 13 sheds light on the country-collaboration map. The outcome corresponds to what was initially expected—the USA is the country that registered the highest number of collaborations, namely 22, with authors from across the world, including Argentina, Austria, Canada, and Italy. As usual, the darker colors in Figure 13 are associated with higher levels of collaboration, while lighter colors with lower levels of inter-country-collaboration.

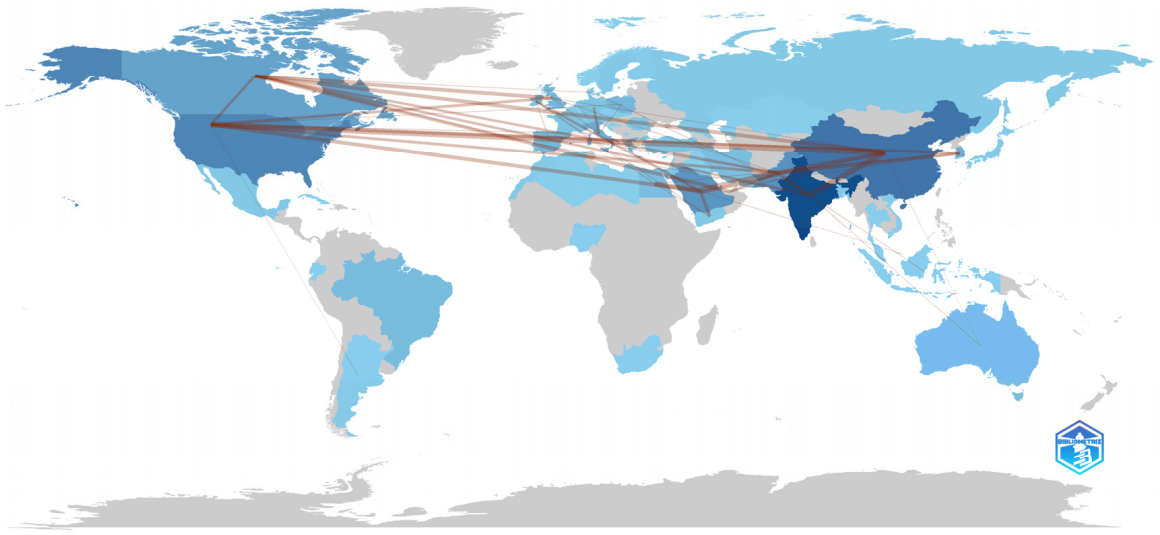


Figure 13. Country-collaboration map.

The top 50 authors and the collaboration network for the papers that are related to the examination of machine learning techniques in fake news research is depicted below in Figure 14.

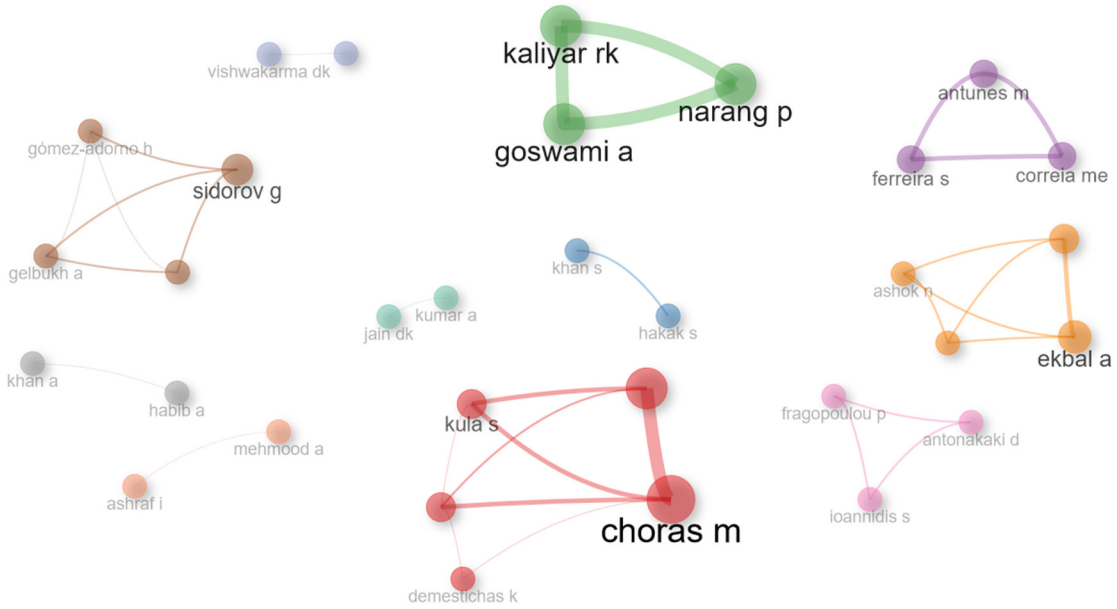


Figure 14. Top 50 authors collaboration network.

3.4. Literature Analysis

In this section, the top 10 most cited documents are extracted from the data collection set and analyzed through multiple perspectives. For each paper, there are details provided regarding the name of the first author, the publishing year, the journal in which it was published, the reference, as well as number of authors, and the region, all meant to emphasize the diversity of the dataset. Furthermore, there are also included some crucial numerical indicators, useful for further analysis: total citations—TC, total citations per year—TYC, and normalized TC—NTC. For more information, kindly consult Table 11.

Table 11. Top 10 most globally cited documents.

No.	Paper (First Author, Year, Journal, Reference)	Number of Authors	Region	Total Citations (TC)	Total Citations per Year (TCY)	Normalized TC (NTC)
1	Bondielli A, 2019, Information Sciences, [46]	2	Italy	208	41.60	6.06
2	Tolosana R, 2020, Information Fusion, [47]	5	Spain	171	42.75	4.84
3	Kaliyar RK, 2021, Multimedia Tools and Applications, [48]	3	India	151	50.33	6.88
4	Sahoo SR, 2021, Applied Soft Computing, [49]	2	India	145	48.33	6.61
5	Hakak S, 2021, Future Generation Computer Systems, [50]	6	Canada, Australia, Pakistan, India, Saudi Arabia	145	48.33	6.61
6	Molina MD, 2021, American Behavioral Scientist, [51]	4	USA	145	48.33	6.61
7	Ahmed H, 2018, Security and Privacy, [52]	3	Canada	143	23.83	2.63
8	Kaliyar RK, 2020, Cognitive Systems Research, [53]	4	India	138	34.50	3.90

Table 11. *Cont.*

No.	Paper (First Author, Year, Journal, Reference)	Number of Authors	Region	Total Citations (TC)	Total Citations per Year (TCY)	Normalized TC (NTC)
9	Kietzmann J, 2020, Business Horizons, [54]	4	Canada, UK, Germany	109	27.25	3.08
10	Küçük D, 2020, ACM Computing Surveys, [55]	2	Turkey	93	23.25	2.63

Additionally, a brief summary is provided for each individual article, including details regarding the topic addressed, the techniques employed, the data used for the analysis, and the goal of the study. All of what follows is intended to inform readers of the substantial body of literature currently available and to point them in the direction of specific works that will be helpful and interesting to their area of focus.

In the last part of the section, we included a word analysis, providing valuable insights regarding the most frequent keywords, bigrams, trigrams, word cloud representations, co-occurrence networks, and thematic maps, all documented and deeply explained.

3.4.1. The Top 10 Most Cited Papers—An Overview

As can be seen from Table 11, the most popular paper in terms of citations, from the extracted data collection set, is the one written by Bondielli et al. [46]. This article has registered an impressive number of total citations (TC), namely 208, but the values obtained for total citations per year (TCY)—41.60 and normalized TC—6.06 are not the highest compared to other values recorded by the other papers included in top 10. Meanwhile, the TCY metric is obtained by dividing the TC by the number of years since the paper's publication, and the NTC metric is determined by dividing the number of citations recorded of a paper by the average number of citations obtained in the same year by all the papers included in the dataset [42]. Thus, in the case of the paper authored by Bondielli et al. [46], the 6.06 value obtained for the NTC has been determined by dividing the 208 citations the paper has received by the average number of citations gathered by the papers in the dataset, namely, 34.32 citations. As a result, it can be stated that the paper written by Bondielli et al. [46] has received 6.06 times more citations than the average of the papers included in the dataset and published in the same year, namely 2019.

The article belonging to Tolosana et al. [47], is ranked in second position, with great values registered for the indicators TC—171, TCY—42.75, and NTC—4.84, followed closely by the third article listed in the table, written by Kaliyar et al. [48], with TC—151, TCY—50.33, and NTC—6.88.

It is also noticeable that the lowest value for TC is 93, for TCY it is 23.25, and for NTC it is 2.63, a fact that highlights the importance, the relevance, and the popularity of these articles in the area of machine learning techniques in fake news research.

By analyzing the number of authors, one can observe that all studies had at least two authors, suggesting that researchers preferred to collaborate in this area. Furthermore, by investigating the "region" column, it can be stated that many of the studies were carried out by teams of authors made up of individuals from different countries, which suggests a high degree of international collaboration in this area.

3.4.2. The Top 10 Most Cited Papers—A Review

In the next pages, each paper listed in the top 10 most globally cited documents is briefly summarized, including the main important details.

The key objectives of the article by Bondielli et al. [46] are to offer reliable methods for identifying false data, decrease the spread of disinformation, examine several approaches for identifying online rumors and fake news, and enhance honesty evaluations in virtual spaces. The comprehensive examination encompassed advanced methods employing

deep learning frameworks, including convolutional neural networks (CNN) and recurrent neural networks (RNN), together with conventional machine learning models, such as logistic regression and decision trees. In addition to the ensemble approaches that are discussed in this article, conditional random field classifiers, random forests, and Hidden Markov models (HMMs) are also covered. The research was conducted using information collected from several web sources including social media platforms like Facebook, Sina Weibo, and Twitter. Depending on the methodology, many studies have claimed accuracy levels ranging from around 0.5 to 0.9, and deep learning approaches have demonstrated promising results, often surpassing the performance of conventional machine learning algorithms. This paper is a valuable work in the scientific community, and the hypothesis was proved also by the significant number of citations, which put it in first position, on the top. The researchers did a great job and provided interesting insights, but nevertheless, a series of issues, such as the demand for wider reference datasets and ongoing research in characteristics that improve detection accuracy, should be further addressed [46].

The article written by Tolosana et al. [47] includes an in-depth analysis of modern facial alteration techniques and detection strategies. The overall key objectives of the research are to improve detection capabilities, provide advances in face manipulation countermeasures, and to offer a thorough knowledge of facial manipulation. Numerous methods are covered in detail, including face synthesis and identity swapping with Generative Adversarial Networks (GANs). With an emphasis on these emerging problems, the main issue that is being addressed is to comprehend and combat the rising threat presented by modified face material. To evaluate the efficacy of detection techniques, the assessment makes use of publicly available databases such as FaceForensics++, DFDC, and Celeb-DF. Among the noteworthy results, there is the ease through which modified information may be detected in controlled environments; but, nonetheless, there are still difficulties in obtaining strong generalizations to real-world variances. In this area, research is still needed to make detection tools more resilient to changing facial manipulation strategies, since detection systems frequently struggle to adjust to new manipulation techniques or databases of the newest generation. However, this paper is truly significant, since it addresses a cutting-edge current subject, and it provides valuable information, but the main weakness in this domain is represented by the fact that few public databases exist for certain forms of modification, such as expression swapping, and in order to further the discipline and investigate deeper, scientists are encouraged to provide more realistic datasets [47].

The academic effort on fake news identification mentioned in the article that belongs to Kaliyar et al. [48] covers the development and evaluation of a model entitled FakeBERT. In this paper, experiments are conducted using pre-trained word embedding techniques (BERT and GloVe) along with the suggested model (FakeBERT) that combines deep learning models (CNN and LSTM). The study distinguishes the performance of these algorithms to define criteria and evaluates their effectiveness using a real-world fake news dataset related to the 2016 U.S. general presidential election. The major objective is to prove that FakeBERT outperforms existing models in terms of accuracy, false positive rate (FPR), false negative rate (FNR), and cross-entropy loss, leading to innovative conclusions. The paper continues to discuss the importance of bi-directional pre-trained word embedding (BERT) for faster training and greater efficiency in fake news classification, and as a result of the investigation, it was highlighted that with an accuracy of 98.90%, the proposed model, FakeBERT, surpasses existing benchmarks. The article provides noteworthy results useful for fake news detection, and is an original work very well conducted by the researchers, its main strength being that it presents to the readers an innovative model, based on the popular BERT, and compares it with other deep learning models using relevant indicators. Anyway, there is still place for improvement and potential research in the future directives, such as in the development of hybrid methodologies for multi-class datasets and the study of echo-chambers in social media for enhanced comprehension of the dissemination of false news [48].

In the next listed article, a combination of a long-term memory (LSTM) deep learning algorithm and several traditional machine learning techniques, including K-Nearest Neighbors, Support Vector Machine, Logistic Regression, Decision Tree, and Naïve Bayes, is used by Sahoo et al. [49] to identify fake news on Facebook through an innovative approach that examines user-generated content and news-related features. The vast dataset includes 42,256,893 posts, 15,328 news stories, and information from over 5000 Facebook profiles, and by considering a wide range of user characteristics and news articles, the primary objective of this research is to increase the accuracy of Facebook's fake news detection system. The outcomes obtained from the study demonstrate that the proposed approach, particularly the LSTM deep learning algorithm, outperforms traditional machine learning algorithms with a remarkable accuracy of 99.4%. Furthermore, the study's ingenious implementation as a Chrome extension for real-time detection highlights the practical utility of the research in fighting erroneous information in online environments, especially on social media. This paper is also considered truly relevant for the academic community according to the number of citations; it provides interesting details, many images, and examples explained in an objective and original manner. Although, there are still some aspects that should be addressed in future work, such as ways to boost the chrome extension's performance, or analyzing the decision making using other deep learning algorithms [49].

The article written by Hakak et al. [50] includes two main datasets, namely Liar, with short, labeled claims, classified as either true or false, concerning a variety of news reports, and ISOT, comprised of 21,417 true news pieces and 23,481 fake news pieces. Fake news is gathered from sources like Politifact and Wikipedia, whereas legitimate information originates from trustworthy sources like Reuters. Via the two datasets, the study explores the detection of fake news using supervised machine learning techniques. The researchers employed Decision Tree, Random Forest, and Extra Tree classifiers hoping to improve accuracy and reduce training durations by using feature extraction. Noise reduction and Named Entity Recognition were applied to the Liar dataset during the feature extraction, which significantly improved classifier performance, presenting higher accuracy, recall, and F1-scores. Similar studies using the ISOT dataset showed that, especially when using the Decision Tree classifier, improved accuracy and shorter training times were achieved following feature extraction. All things considered, the ensemble model with feature extraction produced fantastic outcomes, obtaining 100% accuracy on ISOT, and greatly enhancing the accuracy in Liar. The results obtained in this research are important for future directives related to the classification of fake news; the worked performed is indeed valuable, but in terms of potential areas of improvement in false news detection to a greater extent, there are recommendations to use additional datasets and advanced tuning approaches in future studies [50].

Molina et al. [51] address both theoretical and operational explanations in the paper which serves as an in-depth examination of the concept of fake news. The study combines a mixed-methods approach and a review of the literature to uncover several facets of false news related to message, source, structure, and network dimensions. The authors provide an advanced framework for comprehending the complexity of disinformation by setting out a taxonomy that divides online content into eight categories. The study highlights the potential of machine learning algorithms in detecting false news, while it also underlines the significance of media literacy programs and industry rules. It also identifies elements that may stimulate the development of these technologies, and the work establishes a framework for the further investigation and computational testing of traits proven to be useful in the identification of fake news. Compared to other articles summarized above, one can notice a significative strength of the work conducted in this paper. The article is more focused on the theoretical aspects and possible implications, rather than on the practical side, which helps readers to better understand the concept of "fake news" and to distinguish between features. This may represent a great starting point for future researchers that want to build innovative and efficient models for detecting fake news and combating the spread of this dangerous phenomena [51].

By focusing on false reviews and news, which constitute significant challenges for internet users, the study by Ahmed et al. [52] aims to address the growing issue of fake information. The research provides an original detection model that uses n-gram analysis and highlights feature-extraction approaches. When it comes to identifying opinion spam and fake news, the proposed approach outperforms existing methods. When the model is applied to the news dataset, it achieves an excellent 87% accuracy in distinguishing between fake and real news, while evaluations of the review dataset demonstrate a slight increase (90% accuracy) over previously reported result. The datasets used illustrate the practical applicability of the model, revealing how misleading information influences customer purchasing decisions and shapes public opinion. This impact was especially evident in the context of the 2016 US presidential election. From an objective and critical perspective, the article presents original results and brings to the fore important details, but future research is required for exploring the detection of opinion spam and fake news using multiple methods and features [52].

In order to deal with the pressing issue of false news identification, the study carried out by Kaliyar et al. [53] presents and evaluates a unique deep learning model called FNDNet. Apart from releasing an innovative technique, the investigation aims to progress within the subject by evaluating the model's effectiveness against the state-of-the-art methods now in use. The Kaggle news dataset, which is linked to the 2016 US presidential election, is used to assess these findings. From a personal perspective, this article registered an impressive number of citations, since it addresses an interesting topic in the area of machine learning techniques in fake news research, and intensely examines the differences between numerous deep learning and machine learning algorithms, including a detailed analysis of each and a discussion of the benefits and drawbacks. The research's outcomes demonstrate how effective FNDNet, the suggested deep learning model, is at identifying fake news, being classified as the most accurate model evaluated, with an incredible 98.36% accuracy rate. Regarding the strengths of this work, it is noticed that the study highlights the critical role that state-of-the-art natural language processing techniques—particularly deep learning—have in improving the accuracy of fake news detection, information that is very useful for future research directives [53].

The research by Kietzmann et al. [54] investigates the approach of deepfakes using autoencoder architecture and latent space representation. Key ideas include comprehending approaches, data complexity, and consequences for individuals and groups. The R.E.A.L. risk management framework is its primary proposal. The study conducted in this article is interesting, well organized, provides illustrative examples with explanations, and proves that deepfakes are problematic as they have the possibility to offer personalized entertainment choices but also privacy concerns and threats like online abuse. In order to address deepfake concerns, the report emphasizes the need for powerful partnerships between brands and legislative actions [54].

The study conducted by Küçük et al. [55] offers a comprehensive analysis of stance detection, an essential aspect of natural language processing, with an emphasis on automatically ascertaining an author's stance in relation to a certain target. Position sensing investigations are categorized into several methodologies in the survey, along with applications, tools, and problems. It particularly draws a distinction between programs that use specific position detection modules and those that use general machine learning platforms. The research presented includes anything from opinion surveys to the categorization of rumors and the identification of fake news. In terms of methodology, a variety of machine learning libraries and tools are employed, including scikit-learn, Keras, Theano, and Gensim, as well as SVM in the Weka toolkit. According to the research, position-sensing breakthroughs are facilitated by the development and distribution of source codes. Furthermore, apart from the aforementioned aspects, another strength of the research is that it specifies other domains where location sensing may be applied, including policy discussions, product assessments, and public health monitoring, providing, as well, valuable insights. As for future studies, some notable opportunities consist of context-sensitive tech-

niques, examining post-detection in various contexts and cross-linguistic and multilingual stance detection. Future researchers may start their work by reading this comprehensive article that brings to the fore significative details about this area [55].

Table 12 provides a summary of the works discussed above.

Table 12. Brief summary of the content of top 10 most globally cited documents.

No.	Paper (First Author, Year, Journal, Reference)	Title	Methods Used	Data	Purpose
1	Bondielli A, 2019, Information Sciences, [46]	A survey on fake news and rumour detection techniques	Supervised Learning with Deep Learning (Recurrent Neural Networks and convolutional neural networks). Hybrid Approaches (Combination of RNNs and CNNs). Alternative Approaches (Clustering and Vector Space Models). Tensor Decomposition (CP/PARAFAC Decomposition). Computational-oriented Fact Checking (Knowledge Graphs). Analysis of Diffusion Patterns (Short Diffusion Patterns).	Multiple datasets for examination, including Twitter, Sina Weibo, PHEME, RumorEval, and others.	Examines different techniques and addresses difficulties and upcoming paths to detect false news and rumors in online communication.
2	Tolosana R, 2020, Information Fusion, [47]	Deepfakes and beyond: A Survey of face manipulation and fake detection	Face Synthesis. Identity Swap. Attribute Manipulation. Expression Swapping. Face Morphing. Face De-Identification. Audio-to-Video and Text-to-Video.	Various public databases such as FaceForensics++, DFDC, Celeb-DF, UADFV, and DFFD.	Investigating and comprehending various facial manipulation techniques.
3	Kaliyar RK, 2021, Multimedia Tools and Applications, [48]	FakeBERT: Fake news detection in social media with a BERT-based deep learning approach	Data Collection. Pre-processing. Word Embedding. GloVe Model. BERT Model. The fine-tuning of BERT. Deep Learning Models for Fake News Detection (CNN and LSTM).	The dataset includes 20,800 instances of real-world fake news with respect to the 2016 U.S. general presidential election.	Introducing FakeBERT, a BERT-based model that outperformed other models in identifying fake news.
4	Sahoo SR, 2021, Applied Soft Computing, [49]	Multiple feature-based approach for automatic fake news detection on social networks using deep learning	Feature Extraction. Data Collection and Pre-processing. Machine Learning Algorithms (K-Nearest Neighbors, Support Vector Machine, Logistic Regression, Decision Tree, Naïve Bayes). Deep Learning Algorithm (Long–Short-Term Memory). Performance Evaluation.	The dataset consists of information gathered from 5026 Facebook profiles, a total of 15,328 news articles, and 42,256,893 posts.	Provide a technique for detecting fake news that is especially suited for Facebook and to afterwards put it into practice as a Chrome plugin for real-time detection.
5	Hakak S, 2021, Future Generation Computer Systems, [50]	An ensemble machine learning approach through effective feature extraction to classify fake news	Data Pre-processing. Feature Extraction. Ensemble Model Selection (Random Forest Extra-tree Algorithm Decision Tree). Parameter Tuning (Random Search Hyperparameter Tuning). Supervised Ensemble Approach (Decision Tree, Random Forest, Extra Tree, Bagging)	Liar Dataset: Short, labeled claims, classified as either true or false, concerning a variety of news reports. ISOT Dataset: 21,417 true news pieces and 23,481 fake ones. Fake news is gathered from sources like Politifact and Wikipedia, whereas legitimate information originates from trustworthy sources like Reuters.	Provide an enhanced supervised machine learning model that can be used to identify fake news.

Table 12. *Cont.*

No.	Paper (First Author, Year, Journal, Reference)	Title	Methods Used	Data	Purpose
6	Molina MD, 2021, American Behavioral Scientist, [51]	"Fake News" Is Not Simply False Information: A Concept Explication and Taxonomy of Online Content	Literature review. Content analysis. Taxonomy Development. Machine Learning Algorithm (Potential—Decision Tree model).	Academic articles, trade journals, newspapers, magazines, and other sources connected to the detection and analysis of fake news	Clarifying the concept of fake news, distinguishing characteristics for its recognition, proposing an online content taxonomy, and offering suggestions for the creation of machine learning algorithms designed for false news detection.
7	Ahmed H, 2018, Security and Privacy, [52]	Detecting opinion spam and fake news using text classification	Data Collection. Data Pre-processing. Machine Learning Classifiers (Stochastic Gradient Descent, Support Vector Machine, Linear Support Vector Machine, Logistic Regression, K-Nearest Neighbors, Decision Tree). Evaluation and Analysis.	Dataset 1—Fake Reviews. Dataset 2—Fake News.	Understand and differentiate between authentic and fake content, evaluate the model's capacity to identify false reviews and news by utilizing a variety of classifiers and text processing methods, assess the effect, and examine and contrast the results
8	Kaliyar RK, 2020, Cognitive Systems Research, [53]	FNDNet—A deep convolutional neural network for fake news detection	Machine Learning Models (Multinomial Naive Bayes, K-Nearest Neighbors, Decision Tree, Random Forest). Deep Learning Models (convolutional neural network, Long–Short-Term Memory, Proposed Model—FNDNet). Word Embedding Models (Global Vectors for Word Representation, Word2Vec, Term Frequency-Inverse Document Frequency).	The dataset includes 20,800 instances (Kaggle fake news).	Using the Kaggle news dataset, the article's primary goal is to present and evaluate FNDNet, an innovative deep learning technique for false news detection, along with contrasting its efficacy with current techniques.
9	Kietzmann J, 2020, Business Horizons, [54]	Deepfakes: Trick or treat?	Autoencoder Architecture. Latent Space. Training. Shared Encoder. Manipulation Trick.	Facial trait measures are included in the data. Autoencoders are trained using individual-specific pictures.	To investigate the techniques, information, and consequences of deepfake technology and propose the R.E.A.L. risk management framework.
10	Küçük D, 2020, ACM Computing Surveys, [55]	Stance Detection: A Survey	Stance Detection Techniques (SVM, Naive Bayes, J48, Random Forest, RNN, CNN). Datasets and Annotation. Evaluation Metrics.	Diverse stance datasets including Twitter chats, political conflicts, sports conversations, news headlines, and internet debates in several languages. Datasets made accessible to the public include those from rumor categorization research and those utilized in the Fake News Challenge (FNC).	To enhance the understanding and methods of stance detection in textual content, contribute to applications like sentiment analysis, and offer insights into methods and datasets.

3.4.3. Word Analysis

The keywords, keywords plus, titles, and abstracts included within the dataset will all be carefully examined in the word analysis section, in order to help readers to more deeply comprehend the topic of machine learning techniques in fake news research. This section is expected to contribute to a deeper understanding of the overall discourse by providing insights into recurring themes, highlighting terms that appear on a regular basis, and emphasizing language nuances and particularities, intending to deliver important perspectives on the language landscape of academic works via a systematic examination of word frequencies, associations, and contextual usage.

Table 13 brings to the fore the top 10 most frequently used keywords plus encountered in the selected dataset, which reveal significant themes, such as social media, fake news and information, and discover key focal points and possible trends relevant for the academic literature related to the studied topic of machine learning techniques in fake news research.

Table 13. Top 10 most frequent words in keywords plus.

Words	Occurrences
social media	21
fake news	18
information	17
classification	12
deception	12
misinformation	6
networks	6
news	6
cues	4
diffusion	4

As expected, the leadership position is held by the keyword plus "social media", with 21 occurrences, followed closely by "fake news" with 18 occurrences, "information" with 17 occurrences, "classification" and "deception" each with 12 occurrences, "misinformation", "networks", and "news" each with 6 occurrences, and "cues" and "diffusion", each with 4 occurrences.

Based on the data provided by Table 14, the most frequent words in authors' keywords are listed below, based on the number of occurrences: "fake news"—146 occurrences, "deep learning"—122 occurrences, "machine learning"—99 occurrences, "fake news detection"—74 occurrences, "natural language processing"—62 occurrences, "social media"—54 occurrences, "COVID-19"—29 occurrences, "feature extraction"—27 occurrences, "social networking (online)"—25 occurrences, and "misinformation"—20 occurrences.

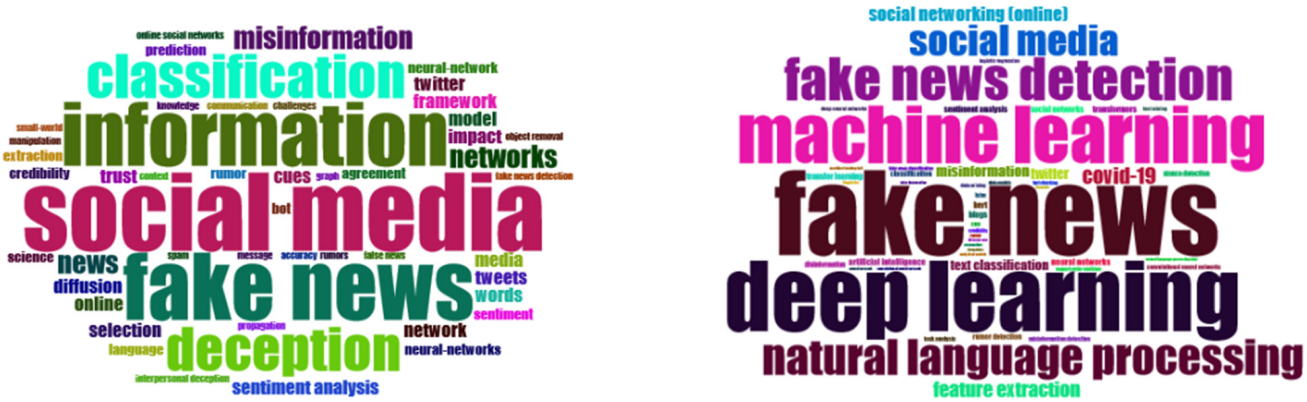
Table 14. Top 10 most frequent words in authors' keywords.

Words	Occurrences
fake news	146
deep learning	122
machine learning	99
fake news detection	74
natural language processing	62
social media	54
COVID-19	29
feature extraction	27
social networking (online)	25
misinformation	20

The most frequently encountered keywords within the collection chosen suggest a trend of implementing advanced computing methods, including natural language pro-

cessing and machine learning techniques, for discovering fake news spread throughout social media platforms. Furthermore, the presence of the word "COVID-19" on this list indicates that researchers were mainly interested in identifying fake news linked to the global pandemic.

Figure 15 depicts the top 50 words, considering the number of occurrences, found in keywords plus and authors' keywords lists, beautifully captured in spectacular pictures with plenty of bursts of color.



(**A**) (**B**)

Figure 15. Top 50 words based on keywords plus (**A**) and authors' keywords (**B**).

Table 15 presents the most frequent bigrams found in abstracts and titles. As expected, there is no surprise that the "fake news" bigram is ranked first for both abstracts and titles, with an impressive value of 994 occurrences in abstracts and 201 in titles. Regarding the second place, the bigram for abstracts is "social media", with 319 occurrences, while for titles it is "news detection", with 108 occurrences. For more details, please consult Table 15.

Table 15. Top 10 most frequent bigrams in abstracts and titles.

Bigrams in Abstracts	Occurrences	Bigrams in Titles	Occurrences
fake news	994	fake news	201
social media	319	news detection	108
machine learning	253	deep learning	49
news detection	217	social media	36
deep learning	180	machine learning	33
learning models	90	news classification	14
neural network	84	learning approach	10
natural language	70	natural language	10
detect fake	63	COVID-fake	9
language processing	61	deep neural	9

Regarding the most frequent trigrams, the first place is occupied by "fake news detection", with 214 occurrences in abstracts and 108 occurrences in titles. In second and the third place for those found in abstracts are "natural language processing", with 61 occurrences and "detect fake news", with 60 occurrences, while, in titles, "fake news classification" has 13 occurrences and "COVID-fake news" has 9 occurrences. Kindly ask you to consult Table 16 for more details.

Table 16. Top 10 most frequent trigrams in abstracts and titles.

Trigrams in Abstracts	Occurrences	Trigrams in Titles	Occurrences
fake news detection	214	fake news detection	108
natural language processing	61	fake news classification	13
detect fake news	60	COVID-fake news	9
social media platforms	51	automatic fake news	8
deep learning models	43	natural language processing	8
detecting fake news	41	Arabic fake news	6
machine learning models	41	deep learning model	6
machine learning algorithms	39	combating fake news	5
support vector machine	29	deep learning approach	5
convolutional neural network	26	deep learning framework	5

With the goal to address and highlight the subject of fake news in a pandemic situation, bigrams and trigrams are revealing fascinating details about themes and trends.

Figure 16 represents a graphical view of the co-occurrence network for the terms in author's keywords.

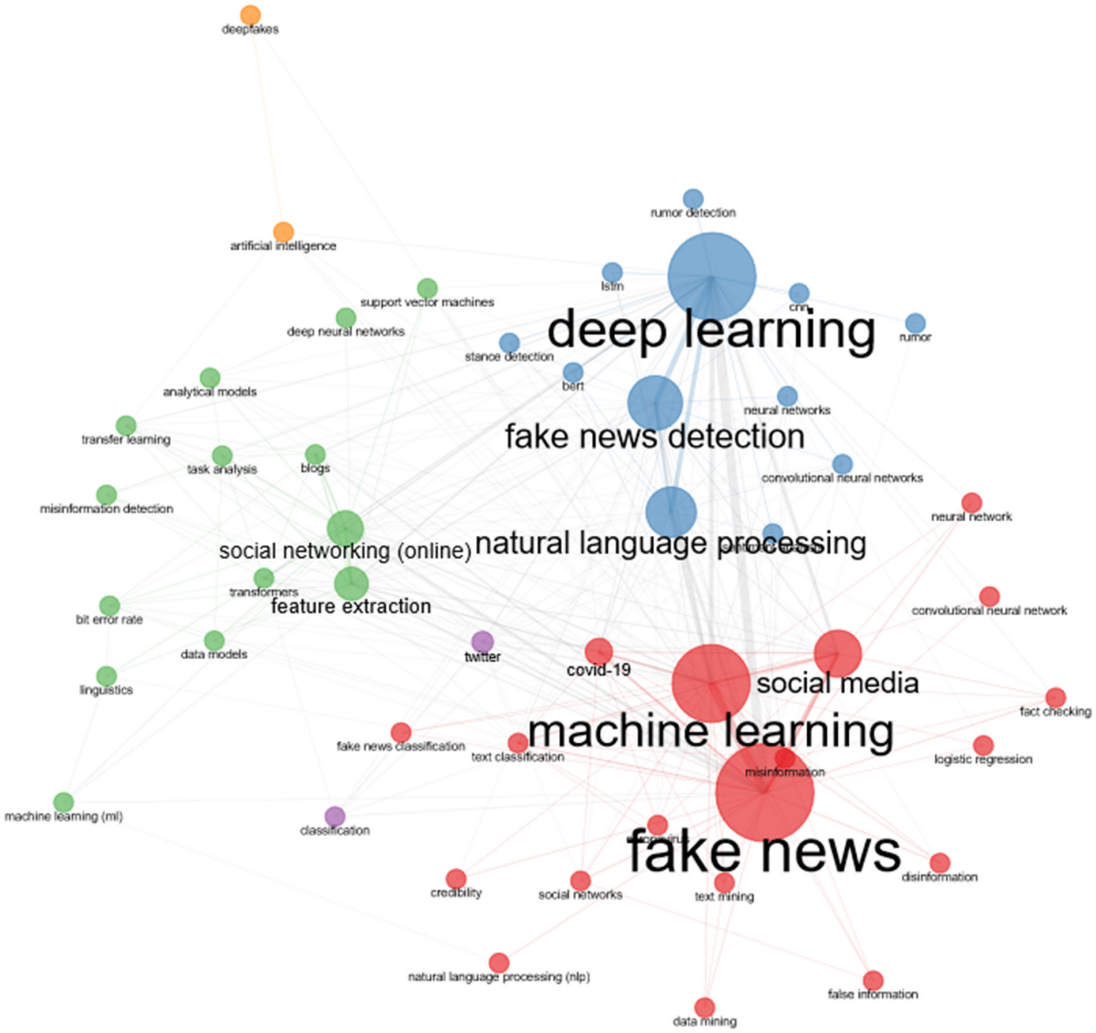


Figure 16. Co-occurrence network for the terms in authors' keywords.

Based on the information in Figure 16, four clusters were delimited:

- Cluster 1—red: fake news, machine learning, social media, COVID-19, misinformation, twitter, text classification, social networks, classification, disinformation, data mining, logistic regression, credibility, fact checking, fake news classification, convolutional neural network, false information, natural language processing (nlp), and neural network;
- Cluster 2—blue: deep learning, fake news detection, natural language processing, neural networks, bert, sentiment analysis, rumor detection, stance detection, cnn, convolutional neural networks, lstm, text mining, rumor, and coronavirus;
- Cluster 3—green: feature extraction, social networking (online), blogs, transfer learning, transformers, misinformation detection, text analysis, deep neural networks, data models, Support Vector Machines, analytical models, bit error rate, linguistics, and machine learning (ml);
- Cluster 4—orange: artificial intelligence and deepfakes.

Figure 17 sheds light on the thematic map based on the authors' keywords, from which four main themes can be outlined: niche themes, motor themes, emerging or declining themes, and basic themes.

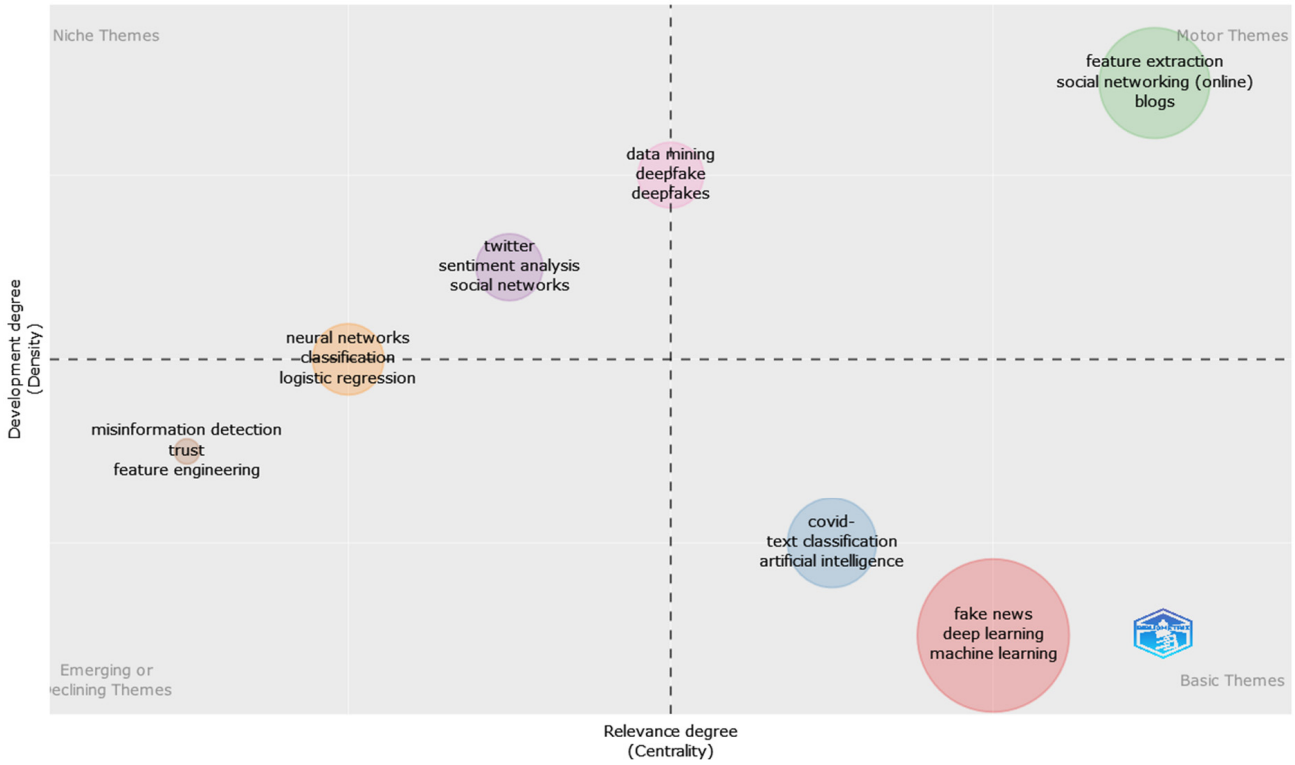


Figure 17. Thematic map based on authors' keywords.

The motor themes include feature extraction, social networking (online), and blogs, along with data mining, deepfake, and deepfakes, which was placed at the border between motor and niche themes. Next on the list are basic themes, which are comprised of fake news, deep learning, machine learning, COVID-19, text classification, and artificial intelligence.

When talking about emerging or declining themes, one can notice misinformation detection, trust, feature engineering, and neural networks, classification, and regression placed at the border with niche themes. Apart from what was already specified, niche themes include twitter, sentiment analysis, and social networks.

By dividing information into distinct thematic groups, this categorization makes it easier for researchers to understand the distribution of the articles and to recognize patterns, trends, and crucial themes in the dataset.

3.5. Mixed Analysis

In this section, a mixed analysis is presented by harnessing the strength of three-fields plots and all of the information gathered up until this point, so as to discover and point out hidden connections between multiple categories, including countries, authors, as well as journals, affiliations, and keywords.

Figure 18 includes a three-fields plot with the first 20 entities found in distinct categories, namely countries (left), authors (middle), and journals (right). The purpose of this is to notice the connection between these three areas, and, based on the information provided here, it can be stated that India holds the leadership position in terms of affiliations for the relevant writers in the studied area: Choras M and Mehmood A are considered the most prominent authors, and the highest number of published articles in the area of machine learning techniques in fake news research is represented by a well-known journal for the scientific community; more specifically, *Neural Computing & Applications.*

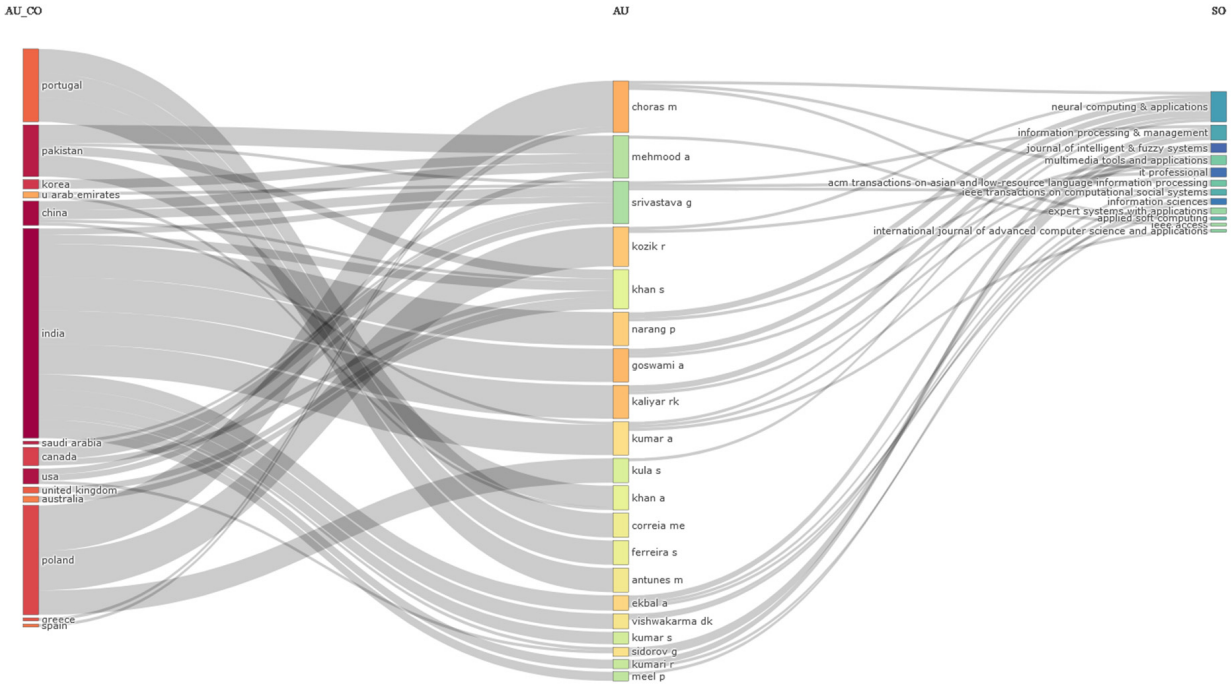


Figure 18. Three-fields plot: countries (**left**), authors (**middle**), journals (**right**).

One can also notice here some insights provided by Figure 18, such as the collaboration between authors from different countries in this area being relatively high and that writers seem to prefer to publish papers in different journals instead of publishing in a single source.

One more three-fields plot representation can be noticed in Figure 19, including three other distinct categories: affiliations (left), authors (middle), and keywords (right). By examining the below picture, it is obvious that all of the listed keywords revolve around very well-defined themes, spotted also above, such as fake news, machine learning techniques and algorithms, social media, and COVID-19. In terms of affiliations, Delhi Technological University seems to occupy the first position at the top.

Moreover, some specific patterns can be noticed here as well. There can be seen authors who are affiliated with multiple universities across the world, both national and international, emphasizing international collaboration, and, on the opposite pole, there are authors who are not affiliated with any of these universities.

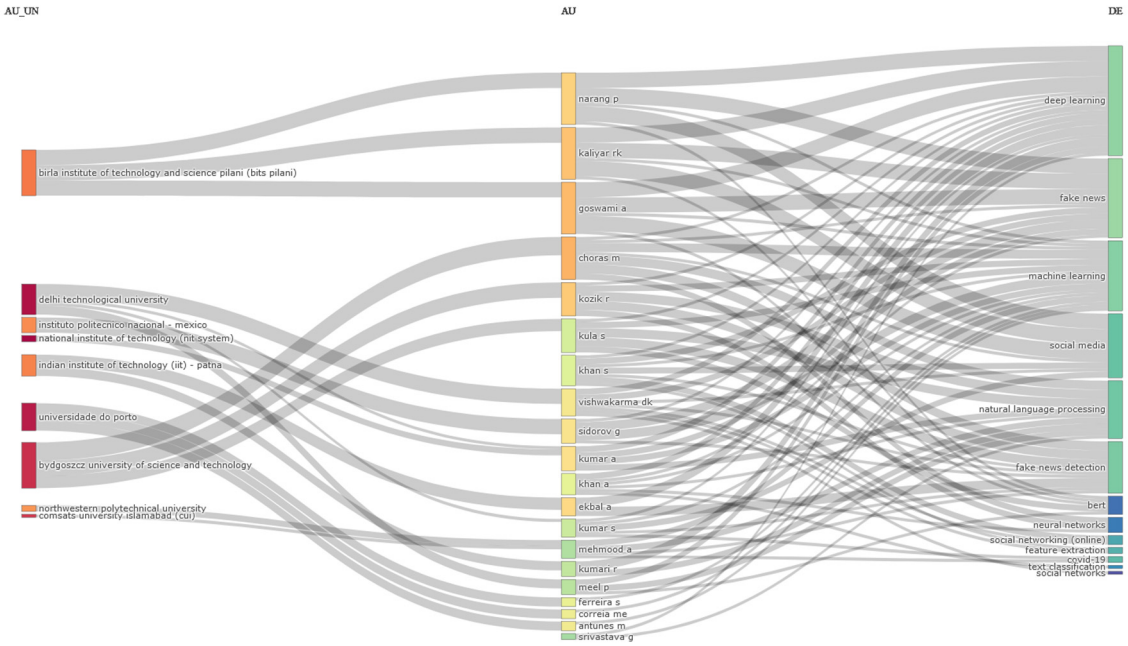


Figure 19. Three-fields plot: affiliations (**left**), authors (**middle**), keywords (**right**).

4. Discussion

This article brought to the surface a detailed bibliometric analysis to uncover trends, perspectives, and insights in the area of machine learning techniques in fake news research, emphasizing the existing scientific literature in this vast field. All articles were extracted and filtered, using the WoS database, resulting in a total number of 346 works marketed as "articles", written in English, and published within the well-defined time period of 2017–2022.

This study includes detailed analyses from different perspectives, such as the most cited articles, the most prolific authors in the field, annual scientific production evolution, a country-collaboration map, collaboration networks, Bradford's law on source clustering, journals' impact based on H-indexes, the most relevant affiliations, scientific production based on country, the most frequent words, bigrams, trigrams, co-occurrence networks, thematic maps, and many more.

A source is considered strengthened when, according to many bibliometric studies, it regularly ranks first for relevance, an aspect which demonstrates both the source's scientific significance and the lasting influence in the academic community. In such cases, the source's fundamental role and outstanding contributions in the field under study is truly obvious, a hypothesis proved by the fact that researchers consistently cite it in their investigations. The same is true for the case of the *IEEE Access* journal, which occupies a significative position in first place, both in the current research article and in other studies such as those carried out in the areas of sentiment analysis with deep learning [56], sentiment analysis in the context of COVID-19 [12], sentiment analysis in the context of COVID-19 vaccines [57], social media research in the age of COVID-19 [58], COVID-19 vaccination misinformation [59], machine learning, and soft computing applications in the textile and clothing supply chain [60].

Another relevant source that was found at the top in this bibliometric analysis is *Expert Systems with Applications*, a journal classified in the forefront position also for other studies in the area of machine learning and artificial intelligence—e.g., machine learning and soft computing applications in the textile and clothing supply chain [60], machine learning in engineering [61], groundbreaking machine learning research across six decades [62], and artificial intelligence applications in supply chain [63].

Regarding the third source mentioned in this paper, the *International Journal of Advanced Computer Science and Applications* has been noted among the top contributors even in other

similar bibliometric works on either opinion mining, sentiment analysis, and emotion understanding [64] or sentiment analysis in times of COVID-19 [12], showing once more the contribution brought by the journal to the research.

Furthermore, an affiliation consistently ranked at the top of numerous bibliometric analyses indicates that its research is of outstanding quality, extremely productive, and of the utmost significance, highlighting its authority as well as significance within the scientific community. In our case, it has been observed that Delhi Technological University has been listed as a top contributor with nine papers. Considering other similar studies focusing on bibliometric analysis in various research areas, it can be observed that Delhi Technological University serves as an example of such a popular affiliation found in these studies, highlighting even more the contribution of the university to the research field. Other bibliometric studies in which Delhi Technological University ranks among the top contributors are in the areas of classification with artificial intelligence and convolutional neural network [65], machine learning used for mental health in social media [66], text mining, and maintenance [67].

Regarding the countries that have been listed as leading contributors, in our case it has been observed that India, China, and the USA are the top contributors. It was observed that these countries are also found in top-contributors lists for other bibliometric studies existing in the academic literature, such as the ones in the areas of opinion mining and sentiment analysis [68], COVID-19 vaccination misinformation [59], social media research in times of COVID-19 [58], social media research in the age of COVID-19 [58], health-related misinformation in social media [69], text mining and maintenance [67], classifications of artificial intelligence using convolutional neural networks [65], sentiment analysis in times of COVID-19 [12], and arrhythmia detection and classification [70]. As a result of these observations, it can be highlighted, once more, the important contribution of these countries to the body of research.

Going back to the listed questions in Section 1, based on the insights discovered during the bibliometric analysis conducted in this paper, there can be some answers provided. Besides the aspects that were already presented above, such as sources, affiliations, and countries, there are still other findings that must be highlighted in this discussion section.

When talking about the cited articles, the first position is held by the paper belonging to Bondielli A, published in *Information Sciences Journal*, in 2019, entitled "A survey on fake news and rumour detection techniques" [46].

Regarding the most prolific authors in this area, Choras M holds the first position on top, with six published articles, followed by Goswami A, Kaliyar RK, Kozik R, Narang P, each with five articles. By paying attention to the number of published articles in the analyzed period, there was observed an increased interest in writing papers in the area of machine learning techniques in fake news research starting with the year 2020, a fact that might be associated with the COVID-19 pandemic.

In terms of collaborations, one can notice that researchers preferred to conduct studies with multiple authors from across the world, rather than individual work, a fact which outlines that the papers written in the area of machine learning techniques in fake news research includes diverse points of view, perspectives, varied resources, and wider audiences, along with cross-cultural insights.

Another question outlined in the introduction was related to insights drawn from the word analysis. By performing a deep investigation of keywords plus, authors' keywords, bigrams, and trigrams, along with providing graphical views of co-occurrence networks and thematic maps, it was observed that the subject of fake news was mostly debated in a COVID-19 pandemic situation, especially based on the social media messages, and the detection of this phenomena was analyzed using multiple machine learning techniques (Support Vector Machine, convolutional neural network, natural language processing. etc.).

That being said, all those findings were deeply explained in the above section, along with providing graphs, images, and tables with all the values and indicators considered. Furthermore, 10 out of the extracted articles, based on the number of citations were pre-

sented and briefly summarized from multiple perspectives, including methods, purpose, number of authors, and data collected. Based on each review, it was observed that for the detection of fake news there were utilized multiple machine learning techniques (Recurrent Neural Networks, convolutional neural networks, Clustering and Vector Space Models, BERT, K-Nearest Neighbors, Support Vector Machine, Logistic Regression, Decision Tree, Naïve Bayes), the datasets were collected from different places (Twitter, FaceForensics++, DFDC, Celeb-DF, Sina Weibo, Facebook, academic articles, trade journals, newspapers, magazines), and most of the articles addressed the issue and the danger of fake news's spread in in multiple areas, including politics and health.

5. Limitations

The initial point that has to be addressed in this section is the fact that the articles were selected from only one source, namely the WoS database. Despite the fact that WoS seems to be one of the most popular databases used in similar researches, covering a wide range of disciplines and journals and being highly recognized by the research community [28–30,33], it should be stated that papers indexed in other databases have been excluded from the current research. Thus, considering other databases for the papers' extraction while using the same keywords and extraction steps, this might have been conducted with a slightly different dataset.

Another essential constraint pertaining to the present research is the selection and utilization of keywords throughout the literature search process. While these keywords were carefully chosen, the inherent issue in this process is the dynamic nature of language and the constantly changing terminology within the area. The exclusion of pertinent research may have resulted from linguistic limitations and the possibility that subtle notions may be represented using other terminology. Moreover, differences in language usage throughout fields or geographical areas may often unintentionally leave out important information.

Additionally, since the research takes language into account, excluding non-English publications might neglect important viewpoints and limit the overall breadth of the present analysis, even if, as Table 2 shows, the number of articles dropped by just three after applying this criterion.

This study's exclusive focus on papers marked as "articles" is another drawback. Though intended, this precision might lead to the exclusion of important details from other kinds of papers, such as books, book chapters, book reviews, or reviews in general. The study findings may be impacted by this choice, as depicted in Table 2.

Furthermore, there is a restriction in the chosen time frame. As the papers written in 2023 have been excluded from the research due to the fact that at the time the dataset was extracted, the 2023 year was not completed yet, the results only refer to the entire timeframe until 2023.

All those predefined filters were carefully picked so as to obtain the final collection of articles, without duplicates or irrelevant papers for the current analysis, focusing both on quality and efficiency, making sure that the extracted dataset will lead to relevant and accurate insights for the research objectives.

Although, the use of a filtered dataset includes some drawbacks that must be stated transparently and objectively. By applying specific rules for extracting the collection of data, in terms of databases, keywords, languages, document types, and time frames, may lead to a limited dataset, some useful works can be omitted or excluded from the analysis, and the results may be slightly affected, since the overall picture of the existing scientific literature may possibly be incomplete.

Regarding the future work directives, interested researchers can try to conduct deeper bibliometric analyses in the area of machine leading techniques in fake news research, using a lager dataset, without applying pre-defined filters during the extraction step (e.g., extending the time frame, using multiple databases).

All the insights presented in this article represent valuable information that can be further used in educational initiatives, ethical implications, collaborations, political and economic decisions, advancements in technology, development, or the improvement of the automated tools for detecting fake news, and many more.

6. Conclusions

Having stated that, taking everything that was described above into consideration, the primary objective of this paper was to present a bibliometric study of the machine learning methods used in fake news research. In the present-day world of technology, fake news is a sensitive topic that may negatively affect a significant amount of the population by manipulating people and inciting fear. Consequently, the rapid spread of false information, especially through the internet, shapes opinions, beliefs, and actions, impacting crucial areas including the economy and global security, political stability, the credibility of institutions, and many other factors.

As a result, after applying certain well-defined criteria, a significant number of 346 articles were retrieved for this bibliometric study, and out of these, the first 10 most cited papers were thoroughly examined and briefly summarized. From these, it was possible to see the variety of techniques employed (BERT, Naïve Bayes, Support Vector Machines, Recurrent Neural Networks, Convolutional Neural Networks, and more), the improved accuracy of the findings, along with the significance, and the achievements of applying machine learning in this area of study.

The most frequent words found in the selected dataset shed light on a variety of topics and subjects which papers explored, such as "machine learning", "fake news detection", "natural language processing", "social media", "COVID-19", "feature extraction", "social networking (online)", and "neural network". It is important to notice here that the COVID-19 outbreak also attracted the interest of many researchers who studied the fake news spread on social media platforms regarding pandemic events.

The platforms from which the datasets for the analyses conducted in the most cited publications were gathered include Twitter, Facebook, Sina Weibo, PHEME, and RumorEval, etc. Moreover, some studies involve public datasets, too.

Regarding the most prolific authors, one can notice the contribution of Choras M, with six published articles, followed closely by other significant authors, alphabetically ordered as Goswami A, Kaliyar RK, Kozik R, and Narang P, each with five articles.

When analyzing the most popular sources preferred for the publication of articles in the area of machine learning techniques in fake news research, the forefront position is held by the *IEEE Access*, with a substantial number of 25 published documents. This conclusion can be drawn also from the H-index analysis.

The Delhi Technological University is ranked in first place for the most relevant affiliation in the studied area, while India, China, and USA are found among the countries marked as the leading contributors.

That being said, the present study revealed essential insights about the area of machine learning techniques in fake news research and addressed the findings, together with both the opportunities and challenges. Future research might focus on related means used in information diffusion, such as misinformation, disinformation, malinformation, deepfakes, rumors, and clickbait. Also, the exploration of new dimensions related to this subject might lead to developing new strategies in combating the spread of fake news and increasing the people's confidence in the news.

Author Contributions: Conceptualization, A.S., I.I., C.D., M.-S.F. and L.-A.C.; Data curation, A.S., I.I., C.D., M.-S.F. and L.-A.C.; Formal analysis, A.S., I.I., C.D., M.-S.F. and L.-A.C.; Investigation, A.S., I.I., C.D., M.-S.F. and L.-A.C.; Methodology, A.S., I.I., C.D., M.-S.F. and L.-A.C.; Resources, A.S.; Software, A.S., I.I., C.D., M.-S.F. and L.-A.C.; Supervision, C.D. and L.-A.C.; Validation, A.S., I.I., C.D., M.-S.F. and L.-A.C.; Visualization, A.S., I.I., C.D. and L.-A.C.; Writing—original draft, A.S., C.D. and L.-A.C.; Writing—review and editing, I.I. and M.-S.F. All authors have read and agreed to the published version of the manuscript.

Funding: The work is supported by a grant of the Romanian Ministry of Research and Innovation, project 42PFE/30.12.2021—'Increasing institutional performance through the development of the infrastructure and research ecosystem of transdisciplinary excellence in the socio-economic field'. This paper was co-financed by The Bucharest University of Economic Studies during the PhD program.

Data Availability Statement: Data are contained within the article.

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

References

1. Siino, M.; Di Nuovo, E.; Tinniriello, I.; La Cascia, M. Fake News Spreaders Detection: Sometimes Attention Is Not All You Need. *Information* **2022**, *13*, 426. [CrossRef]
2. Dennis, A.R.; Galletta, D.F.; Webster, J. Special Issue: Fake News on the Internet. *J. Manag. Inf. Syst.* **2021**, *38*, 893–897. [CrossRef]
3. La, T.-V.; Dao, M.-S.; Le, D.-D.; Thai, K.-P.; Nguyen, Q.-H.; Phan-Thi, T.-K. Leverage Boosting and Transformer on Text-Image Matching for Cheap Fakes Detection. *Algorithms* **2022**, *15*, 423. [CrossRef]
4. Carmi, E.; Yates, S.J.; Lockley, E.; Pawluczuk, A. Data Citizenship: Rethinking Data Literacy in the Age of Disinformation, Misinformation, and Malinformation. *Internet Policy Rev.* **2020**, *9*, 1–22. [CrossRef]
5. Gradoń, K.T.; Hołyst, J.A.; Moy, W.R.; Sienkiewicz, J.; Suchecki, K. Countering Misinformation: A Multidisciplinary Approach. *Big Data Soc.* **2021**, *8*, 205395172110138. [CrossRef]
6. Wardle, C. *Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making (2017)*; Council of Europe: Strasbourg, France, 2017.
7. Taylor and Francis Website Misinformation vs Disinformation-Taylor & Francis Insights. Available online: https://insights.taylorandfrancis.com/social-justice/misinformation-vs-disinformation/ (accessed on 9 December 2023).
8. Lazer, D.M.J.; Baum, M.A.; Benkler, Y.; Berinsky, A.J.; Greenhill, K.M.; Menczer, F.; Metzger, M.J.; Nyhan, B.; Pennycook, G.; Rothschild, D.; et al. The Science of Fake News. *Science* **2018**, *359*, 1094–1096. [CrossRef] [PubMed]
9. Adjin-Tettey, T.D. Combating Fake News, Disinformation, and Misinformation: Experimental Evidence for Media Literacy Education. *Cogent Arts Humanit.* **2022**, *9*, 2037229. [CrossRef]
10. Aïmeur, E.; Amri, S.; Brassard, G. Fake News, Disinformation and Misinformation in Social Media: A Review. *Soc. Netw. Anal. Min.* **2023**, *13*, 30. [CrossRef]
11. Sadiku, M.N.O.; Eze, T.P.; Musa, S.M. Fake news and misinformation. *Int. J. Adv. Sci. Res. Eng.* **2018**, *4*, 563–576. [CrossRef]
12. Sandu, A.; Cotfas, L.-A.; Delcea, C.; Crăciun, L.; Molanescu, A.G. Sentiment Analysis in the Age of COVID-19: A Bibliometric Perspective. *Information* **2023**, *14*, 659. [CrossRef]
13. Alarfaj, F.K.; Khan, J.A. Deep Dive into Fake News Detection: Feature-Centric Classification with Ensemble and Deep Learning Methods. *Algorithms* **2023**, *16*, 507. [CrossRef]
14. Kasnesis, P.; Toumanidis, L.; Patrikakis, C. Combating Fake News with Transformers: A Comparative Analysis of Stance Detection and Subjectivity Analysis. *Information* **2021**, *12*, 409. [CrossRef]
15. Leonardi, S.; Rizzo, G.; Morisio, M. Automated Classification of Fake News Spreaders to Break the Misinformation Chain. *Information* **2021**, *12*, 248. [CrossRef]
16. Alghamdi, J.; Lin, Y.; Luo, S. A Comparative Study of Machine Learning and Deep Learning Techniques for Fake News Detection. *Information* **2022**, *13*, 576. [CrossRef]
17. Buzea, M.C.; Trausan-Matu, S.; Rebedea, T. Automatic Fake News Detection for Romanian Online News. *Information* **2022**, *13*, 151. [CrossRef]
18. Castiello, M.; Conte, D.; Iscaro, S. Using Epidemiological Models to Predict the Spread of Information on Twitter. *Algorithms* **2023**, *16*, 391. [CrossRef]
19. Oliveira, N.; Pisa, P.T.; Lopez, M.A.; Medeiros, D.; Mattos, D. Identifying Fake News on Social Networks Based on Natural Language Processing: Trends and Challenges. *Information* **2021**, *12*, 38. [CrossRef]
20. Cotfas, L.-A.; Delcea, C.; Roxin, I.; Ioanas, C.; Gherai, D.S.; Tajariol, F. The Longest Month: Analyzing COVID-19 Vaccination Opinions Dynamics from Tweets in the Month Following the First Vaccine Announcement. *IEEE Access* **2021**, *9*, 33203–33223. [CrossRef]
21. Delcea, C.; Cotfas, L.-A.; Crăciun, L.; Molănescu, A.G. New Wave of COVID-19 Vaccine Opinions in the Month the 3rd Booster Dose Arrived. *Vaccines* **2022**, *10*, 881. [CrossRef]
22. van der Linden, S.; Roozenbeek, J.; Compton, J. Inoculating Against Fake News About COVID-19. *Front. Psychol.* **2020**, *11*, 2928. [CrossRef]
23. De Magistris, G.; Russo, S.; Roma, P.; Starczewski, J.; Napoli, C. An Explainable Fake News Detector Based on Named Entity Recognition and Stance Classification Applied to COVID-19. *Information* **2022**, *13*, 137. [CrossRef]
24. Apuke, O.D.; Omar, B. Fake News and COVID-19: Modelling the Predictors of Fake News Sharing among Social Media Users. *Telemat. Inform.* **2021**, *56*, 101475. [CrossRef] [PubMed]
25. Ceron, W.; de-Lima-Santos, M.-F.; Quiles, M.G. Fake News Agenda in the Era of COVID-19: Identifying Trends through Fact-Checking Content. *Online Soc. Netw. Media* **2021**, *21*, 100116. [CrossRef]
26. WoS Web of Science. Available online: https://www.webofscience.com (accessed on 9 September 2023).

27. Liu, F. Retrieval Strategy and Possible Explanations for the Abnormal Growth of Research Publications: Re-Evaluating a Bibliometric Analysis of Climate Change. *Scientometrics* **2022**, *128*, 853–859. [CrossRef] [PubMed]
28. Bakır, M.; Özdemir, E.; Akan, Ş.; Atalık, Ö. A Bibliometric Analysis of Airport Service Quality. *J. Air Transp. Manag.* **2022**, *104*, 102273. [CrossRef]
29. Cobo, M.J.; Martínez, M.A.; Gutiérrez-Salcedo, M.; Fujita, H.; Herrera-Viedma, E. 25 Years at Knowledge-Based Systems: A Bibliometric Analysis. *Knowl. Based Syst.* **2015**, *80*, 3–13. [CrossRef]
30. Modak, N.M.; Merigó, J.M.; Weber, R.; Manzor, F.; Ortúzar, J.D.D. Fifty Years of Transportation Research Journals: A Bibliometric Overview. *Transp. Res. Part Policy Pract.* **2019**, *120*, 188–223. [CrossRef]
31. Sandu, A.; Ioanas, I.; Delcea, C.; Geanta, L.-M.; Cotfas, L.-A. Mapping the Landscape of Misinformation Detection: A Bibliometric Approach. *Information* **2024**, *15*, 60. [CrossRef]
32. Liu, W. The Data Source of This Study Is Web of Science Core Collection? Not Enough. *Scientometrics* **2019**, *121*, 1815–1824. [CrossRef]
33. Mulet-Forteza, C.; Martorell-Cunill, O.; Merigó, J.M.; Genovart-Balaguer, J.; Mauleon-Mendez, E. Twenty Five Years of the Journal of Travel & Tourism Marketing: A Bibliometric Ranking. *J. Travel Tour. Mark.* **2018**, *35*, 1201–1221. [CrossRef]
34. Marín-Rodríguez, N.J.; González-Ruiz, J.D.; Valencia-Arias, A. Incorporating Green Bonds into Portfolio Investments: Recent Trends and Further Research. *Sustainability* **2023**, *15*, 14897. [CrossRef]
35. Stefanis, C.; Giorgi, E.; Tselemponis, G.; Voidarou, C.; Skoufos, I.; Tzora, A.; Tsigalou, C.; Kourkoutas, Y.; Constantinidis, T.C.; Bezirtzoglou, E. Terroir in View of Bibliometrics. *Stats* **2023**, *6*, 956–979. [CrossRef]
36. Gorski, A.-T.; Ranf, E.-D.; Badea, D.; Halmaghi, E.-E.; Gorski, H. Education for Sustainability—Some Bibliometric Insights. *Sustainability* **2023**, *15*, 14916. [CrossRef]
37. Fatma, N.; Haleem, A. Exploring the Nexus of Eco-Innovation and Sustainable Development: A Bibliometric Review and Analysis. *Sustainability* **2023**, *15*, 12281. [CrossRef]
38. WoS Document Types. Available online: https://webofscience.help.clarivate.com/en-us/Content/document-types.html (accessed on 3 December 2023).
39. Donner, P. Document Type Assignment Accuracy in the Journal Citation Index Data of Web of Science. *Scientometrics* **2017**, *113*, 219–236. [CrossRef]
40. Aria, M.; Cuccurullo, C. Bibliometrix: An R-Tool for Comprehensive Science Mapping Analysis. *J. Informetr.* **2017**, *11*, 959–975. [CrossRef]
41. Domenteanu, A.; Delcea, C.; Chiriță, N.; Ioanăș, C. From Data to Insights: A Bibliometric Assessment of Agent-Based Modeling Applications in Transportation. *Appl. Sci.* **2023**, *13*, 12693. [CrossRef]
42. Delcea, C.; Javed, S.A.; Florescu, M.-S.; Ioanas, C.; Cotfas, L.-A. 35 Years of Grey System Theory in Economics and Education. *Kybernetes ahead-of-print*. **2023**. [CrossRef]
43. Cibu, B.; Delcea, C.; Domenteanu, A.; Dumitrescu, G. Mapping the Evolution of Cybernetics: A Bibliometric Perspective. *Computers* **2023**, *12*, 237. [CrossRef]
44. Wardikar, V. Application of Bradford's Law of Scattering to the Literature of Library & Information Science: A Study of Doctoral Theses Citations Submitted to the Universities of Maharashtra, India. *Libr. Philos. Pract. E J.* **2023**, *1054*, 1–45.
45. RDRR Website Bradford: Bradford's Law in Bibliometrix: Comprehensive Science Mapping Analysis. Available online: https://rdrr.io/cran/bibliometrix/man/bradford.html (accessed on 21 November 2023).
46. Bondielli, A.; Marcelloni, F. A Survey on Fake News and Rumour Detection Techniques. *Inf. Sci.* **2019**, *497*, 38–55. [CrossRef]
47. Tolosana, R.; Vera-Rodriguez, R.; Fierrez, J.; Morales, A.; Ortega-Garcia, J. Deepfakes and beyond: A Survey of Face Manipulation and Fake Detection. *Inf. Fusion* **2020**, *64*, 131–148. [CrossRef]
48. Kaliyar, R.K.; Goswami, A.; Narang, P. FakeBERT: Fake News Detection in Social Media with a BERT-Based Deep Learning Approach. *Multimed. Tools Appl.* **2021**, *80*, 11765–11788. [CrossRef] [PubMed]
49. Sahoo, S.R.; Gupta, B.B. Multiple Features Based Approach for Automatic Fake News Detection on Social Networks Using Deep Learning. *Appl. Soft Comput.* **2021**, *100*, 106983. [CrossRef]
50. Hakak, S.; Alazab, M.; Khan, S.; Gadekallu, T.R.; Maddikunta, P.K.R.; Khan, W.Z. An Ensemble Machine Learning Approach through Effective Feature Extraction to Classify Fake News. *Future Gener. Comput. Syst.* **2021**, *117*, 47–58. [CrossRef]
51. Molina, M.D.; Sundar, S.S.; Le, T.; Lee, D. "Fake News" Is Not Simply False Information: A Concept Explication and Taxonomy of Online Content. *Am. Behav. Sci.* **2019**, *65*, 180–212. [CrossRef]
52. Ahmed, H.; Traore, I.; Saad, S. Detecting Opinion Spams and Fake News Using Text Classification. *Secur. Priv.* **2017**, *1*, e9. [CrossRef]
53. Kaliyar, R.K.; Goswami, A.; Narang, P.; Sinha, S. FNDNet—A Deep Convolutional Neural Network for Fake News Detection. *Cogn. Syst. Res.* **2020**, *61*, 32–44. [CrossRef]
54. Kietzmann, J.; Lee, L.W.; McCarthy, I.P.; Kietzmann, T.C. Deepfakes: Trick or Treat? *Bus. Horiz.* **2020**, *63*, 135–146. [CrossRef]
55. Küçük, D.; Can, F. Stance Detection: A Survey. *ACM Comput. Surv.* **2020**, *53*, 1–37. [CrossRef]
56. Puteh, N.; Ali bin Saip, M.; Zabidin Husin, M.; Hussain, A. Sentiment Analysis with Deep Learning: A Bibliometric Review. *Turk. J. Comput. Math. Educ. (TURCOMAT)* **2021**, *12*, 1509–1519.
57. Sarirete, A. A Bibliometric Analysis of COVID-19 Vaccines and Sentiment Analysis. *Procedia Comput. Sci.* **2021**, *194*, 280–287. [CrossRef] [PubMed]

58. Michailidis, P.D. Visualizing Social Media Research in the Age of COVID-19. *Information* **2022**, *13*, 372. [CrossRef]
59. Mahajan, R.; Gupta, P. A Bibliometric Analysis on the Dissemination of COVID-19 Vaccine Misinformation on Social Media. *J. Content Community Commun.* **2021**, *14*, 218–229. [CrossRef]
60. Arora, S.; Majumdar, A. Machine Learning and Soft Computing Applications in Textile and Clothing Supply Chain: Bibliometric and Network Analyses to Delineate Future Research Agenda. *Expert Syst. Appl.* **2022**, *200*, 117000. [CrossRef]
61. Su, M.; Peng, H.; Li, S. A Visualized Bibliometric Analysis of Mapping Research Trends of Machine Learning in Engineering (MLE). *Expert Syst. Appl.* **2021**, *186*, 115728. [CrossRef]
62. Ezugwu, A.E.; Greeff, J.; Ho, Y.-S. A Comprehensive Study of Groundbreaking Machine Learning Research: Analyzing Highly Cited and Impactful Publications across Six Decades. *J. Eng. Res.* **2023**, S2307187723002882. [CrossRef]
63. Riahi, Y.; Saikouk, T.; Gunasekaran, A.; Badraoui, I. Artificial Intelligence Applications in Supply Chain: A Descriptive Bibliometric Analysis and Future Research Directions. *Expert Syst. Appl.* **2021**, *173*, 114702. [CrossRef]
64. Sanchez-Nunez, P.; Cobo, M.J.; Heras-Pedrosa, C.D.L.; Pelaez, J.I.; Herrera-Viedma, E. Opinion Mining, Sentiment Analysis and Emotion Understanding in Advertising: A Bibliometric Analysis. *IEEE Access* **2020**, *8*, 134563–134576. [CrossRef]
65. Kumar, S.; Patil, R.; Kumawat, V.; Rai, Y.; Krishnan, N.; Singh, S. A Bibliometric Analysis of Plant Disease Classification with Artificial Intelligence Using Convolutional Neural Network. *Libr. Philos. Pract.* **2021**, *2021*, 1–14.
66. Kim, J.; Lee, D.; Park, E. Machine Learning for Mental Health in Social Media: Bibliometric Study. *J. Med. Internet Res.* **2021**, *23*, e24870. [CrossRef] [PubMed]
67. Gonçalves Júnior, E.; Gonçalves, V.; Carvalho, Á. Bibliometric Study in Text Mining and Maintenance. *Int. J. Sci. Res. IJSR* **2018**, *7*, 1796–1801. [CrossRef]
68. Musa, I.H.; Zamit, I.; Xu, K.; Boutouhami, K.; Qi, G. A Comprehensive Bibliometric Analysis on Opinion Mining and Sentiment Analysis Global Research Output. *J. Inf. Sci.* **2023**, *49*, 1506–1516. [CrossRef]
69. Yeung, A.W.K.; Tosevska, A.; Klager, E.; Eibensteiner, F.; Tsagkaris, C.; Parvanov, E.D.; Nawaz, F.A.; Völkl-Kernstock, S.; Schaden, E.; Kletecka-Pulker, M.; et al. Medical and Health-Related Misinformation on Social Media: Bibliometric Study of the Scientific Literature. *J. Med. Internet Res.* **2022**, *24*, e28152. [CrossRef]
70. Gronthy, U.U.; Biswas, U.; Tapu, S.; Samad, M.A.; Nahid, A.-A. A Bibliometric Analysis on Arrhythmia Detection and Classification from 2005 to 2022. *Diagnostics* **2023**, *13*, 1732. [CrossRef]