technologies

Review

Fake News Detection Revisited: An Extensive Review of Theoretical Frameworks, Dataset Assessments, Model Constraints, and Forward-Looking Research Agendas

Sheetal Harris [1,†], Hassan Jalil Hadi [1,2,*,†], Naveed Ahmad [2] and Mohammed Ali Alshara [2,3]

[1] School of Cyber Science and Engineering, Wuhan University, Wuhan 430072, China; sheetal.harris@whu.edu.cn
[2] Prince Sultan University, Riyadh 12435, Saudi Arabia; nahmed@psu.edu.sa (N.A.); malshara@psu.edu.sa (M.A.A.)
[3] College of Computer Sciences and Information for Educational and Quality Affairs, Al-Imam Muhammad Ibn Saud Islamic University, Riyadh 12435, Saudi Arabia
* Correspondence: hjhwhu@whu.edu.cn
† These authors contributed equally to this work.

Abstract: The emergence and acceptance of digital technology have caused information pollution and an infodemic on Online Social Networks (OSNs), blogs, and online websites. The malicious broadcast of illegal, objectionable and misleading content causes behavioural changes and social unrest, impacts economic growth and national security, and threatens users' safety. The proliferation of AI-generated misleading content has further intensified the current situation. In the previous literature, state-of-the-art (SOTA) methods have been implemented for Fake News Detection (FND). However, the existing research lacks multidisciplinary considerations for FND based on theories on FN and OSN users. Theories' analysis provides insights into effective and automated detection mechanisms for FN, and the intentions and causes behind wide-scale FN propagation. This review evaluates the available datasets, FND techniques, and approaches and their limitations. The novel contribution of this review is the analysis of the FND in linguistics, healthcare, communication, and other related fields. It also summarises the explicable methods for FN dissemination, identification and mitigation. The research identifies that the prediction performance of pre-trained transformer models provides fresh impetus for multilingual (even for resource-constrained languages), multidomain, and multimodal FND. Their limits and prediction capabilities must be harnessed further to combat FN. It is possible by large-sized, multidomain, multimodal, cross-lingual, multilingual, labelled and unlabelled dataset curation and implementation. SOTA Large Language Models (LLMs) are the innovation, and their strengths should be focused on and researched to combat FN, deepfakes, and AI-generated content on OSNs and online sources. The study highlights the significance of human cognitive abilities and the potential of AI in the domain of FND. Finally, we suggest promising future research directions for FND and mitigation.

Keywords: fake news detection; dataset evaluation; machine learning; deep learning; natural language processing; social networks

check for updates

Citation: Harris, S.; Hadi, H.J.; Ahmad, N.; Alshara, M.A. Fake News Detection Revisited: An Extensive Review of Theoretical Frameworks, Dataset Assessments, Model Constraints, and Forward-Looking Research Agendas. *Technologies* **2024**, *12*, 222. https://doi.org/10.3390/technologies12110222

Academic Editor: Lipo Wang

Received: 2 September 2024
Revised: 4 October 2024
Accepted: 16 October 2024
Published: 6 November 2024

1. Introduction

News portals, Online Social Networks (OSNs), and search engines have become indispensable sources of information for users. The convenience, wide circulation, and inexpensive nature of online portals and OSNs have persuaded numerous users to switch from traditional media [1]. Unverified content is also published to draw users' attention and create traffic to online portals. News consumption through online news portals and OSNs has increased manifold compared to traditional mainstream media [2]. The decentralised nature of online media sources and the absence of censorship, accountability and authority to validate facts and figures has increased fake news (FN) [3]. The users are enticed by clickbait trigrams with less effort. The appearance of FN in public discourse has intensified since 2016, the UK Brexit referendum, and the United States presidential election. Ironically, fake election stories have garnered more shares, reactions, and comments on Facebook (8,711,000) than the most-discussed election reports from major news websites (7,367,000). The explosion of misinformation and disinformation in the digital environment during the pandemic was so intensive that it was termed an "infodemic". FN propagates with higher velocity and greater impact rather than true news stories.

Additionally, the AI-generated FN images have added fuel to the fire. AI-generated FN, forged images, news with counterfeit facts, and satire appear authentic [4,5]. Thus, users are unable to identify authentic news. Users share biased news, intentionally and unintentionally, that is accessible to other users. This process continues as a chain reaction. Lastly, this chain reaction leads to the manipulation of public opinion. The other major factor is the absence or lack of censorship [2], which causes forged news to spread widely and robustly. Due to the echo chamber effect on search engines and OSNs, biased information and distorted facts are abundant and fraught with misinformation and disinformation. FN is a false and fabricated narrative that manipulates the reader's mindset, behaviour, emotions, and reactions worldwide [6]. FN promotes rumours, religious bias, hidden monetary pursuits, and political agendas [7]. FN targeted towards a targeted ethnicity, race, religion and political affiliations are spread to provoke violence and turmoil in society [8]. Thus, it results in psychological pressures and physical damage, immensely impacts social stability, and undermines democracy. FN is also used as a state-of-the-art (SOTA) information warfare tool and tactic to spread negative propaganda [9], conspiracy, and misconception about a state or country worldwide. It is essential to detect and report FN in terms of style, content, and social context [10]. The process of Fake News Detection (FND) depends on identifying the psychological and social features [11] that instigate a user to believe and share FN. The users are incompetent and technologically illiterate to differentiate between fake and credible news or to follow fact-checking websites [12]. The incoherent and inconsistent approach of fact-checking websites is a disincentive and a source of misperception. Different portals share diverse views about the same FN, and this results in confusion among the users [13]. The interdisciplinary studies and theories have demonstrated that the human ability to identify FN and deception is execrable [6]. Various theories confirm that established beliefs [14], biases and approaches [15,16], frequent exposure to FN [17], and, often, peer pressure lead users to believe FN [18] and direct them to FN dissemination [19]. FND is of great eminence due to the gravity of the situation and its impact on users, societies, governments, and countries [8,20]. The unique characteristics of OSNs and online sources propagate yellow journalism with vast distribution and in a cycle format [21,22]. There are dedicated profiles on the OSNs (Twitter, Facebook, WhatsApp, etc.) to disseminate FN [23]. Various SOTA techniques, models, and approaches have been studied for FND. They inspire researchers and is a way forward to prevent worldwide misadventures, i.e., information warfare.

FND, prevention, and intervention are challenging. The research shows the limitations of traditional FND algorithms, in which the supporting attributes also have restrictions because of loads of unstructured data and limited-size datasets [24]. Researchers have used several knowledge-based and deep-structured learning mechanisms to identify the creators of FN [25]. The fact remains that humans are more responsible for sharing misleading information than bots or cyborgs [26,27], with irreversible and unprecedented effects overall. Dataset bias and topic restriction [28] are the prime hindrances since the existing literature is focused on different research directions and scopes. A humongous range of datasets has been used for FND. However, the lack of cross-domain, multidomain, and multilingual datasets hinders FND thus far. Lack of interpretability and hyperparameter optimisation [29] are the main challenges that encumber its wide acceptance. The study of FND on online sources and OSNs with the help of SOTA techniques [28] to evaluate different datasets [30] is apposite. Various DL and ML approaches have been used [31] and demonstrate promising results for FND. Previous research shows that machine learning

(ML) [32,33], deep learning (DL) [34], natural language processing (NLP) [35,36], and knowledge-based techniques [37] are some of the most adopted methods for FND using a range of datasets. The existing literature also uses sentiment analysis [38] and information retrieval [12] to address the issue proactively. Feature-based research (linguistic features, statistical-text features, temporal–structural features, and hybrid features) have also been conducted [39–41]. Consequently, it is essential to study interdisciplinary theories to identify the cause of FN transmission, consumption, and propagation. With the appropriate background knowledge about human behaviour and the psychology of FN, the focus on available datasets and FND approaches can ensure a trustworthy mechanism for FND.

1.1. Fake News (FN)

Fake news (FN) has remained a debatable and disagreed term, to date. FN is defined by the Cambridge Dictionary as deceptive news reports circulated online or through other media that are produced to influence political opinions or serve as humour. The Oxford Dictionary terms FN as false information about incidents published and accessed on websites. The Merriam-Webster dictionary explains FN as false and fabricated material reported in publications, e.g., newspapers, news magazines, or television news programs. Hence, FN is false, counterfeit, and forged information circulated in traditional and digital (social and online) media [42]. Interestingly, the definitions and their explanations determine that FN is created and shared to mislead and stimulate public opinion. This crux of the problem relates to FN as an intentional or unintentional attack on users to shape their conceptions and perspectives to promote their hidden agenda and propaganda. The existing literature reports FN as non-factual and fabricated content. However, the inclusion and exclusion of several related ideas, such as satire, hoaxes, rumours, conspiracy theories, and misinformation from the defined term, is a point of contention. Furthermore, many phrases and ideas in the literature are correlated to FN [43]. The researchers have also used correlated terms for FN, such as false information [21], deceptive news [44], misinformation [45,46], and disinformation [47]. In the current scenario of widespread FN threat, FN has also been defined as information pollution [48], information disorder [46,49], and information warfare [50,51]. Various types of FN with respect to writing style, propagation impact and velocity are shown in Figure 1. It demonstrates that with an increase in the severity of the fabricated material, its propagation velocity and intensity of adverse effects escalate. The impact and velocity ratio also validates that FN promulgates more than factual news.

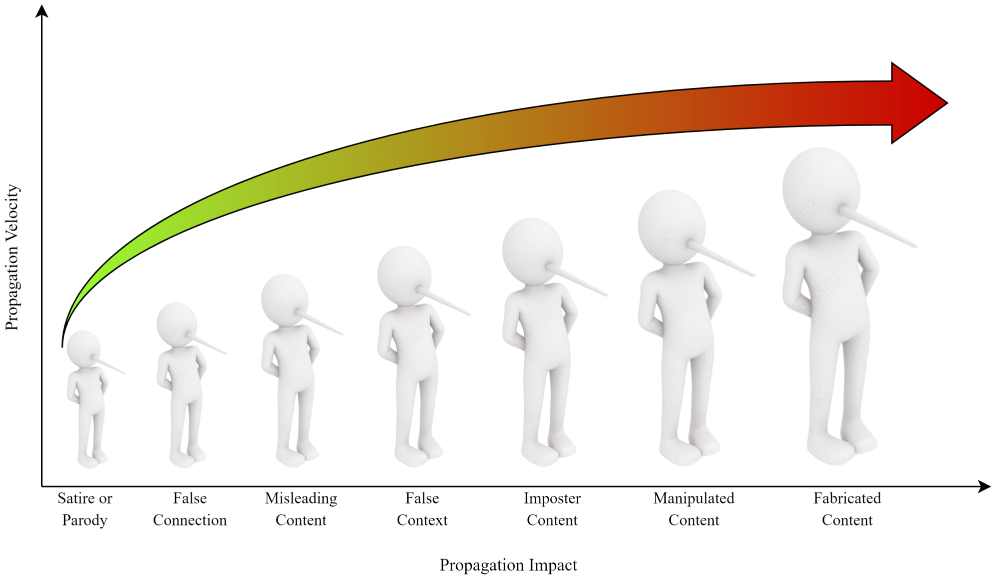

Figure 1. FN types in terms of veracity value and velocity [52].

Analysis of the burgeoning literature determines that the agreed terms on FN are based on two common essential characteristics, i.e., the intention and authenticity of the news content. The purpose or motivation refers to the intent of spreading FN. The objective is either to deceive, misguide, and mislead the readers or to distress, create unrest, and harm them. The intention factor also explains why deceptive news influences more users than satirical news. The chief reason is that such misleading material is specifically written and designed to persuade (mislead) the targeted audience. AI-generated FN has further added variety to the deception. Therefore, another aspect of widespread misinformation is added in [53] to explain the current state of FN as shown in Figure 2. Moreover, malicious users propagate deceptive news with ill intentions and for monetary benefits to augment its impact on society. Therefore, once the intentions behind non-factual information are determined, intervention strategies can be more appropriate and effective. Secondly, the truthfulness and legitimacy of the content indicate if the disseminated news or content on online portals and OSNs can be verified and is authentic. Therefore, FN is termed false information, misinformation and disinformation.

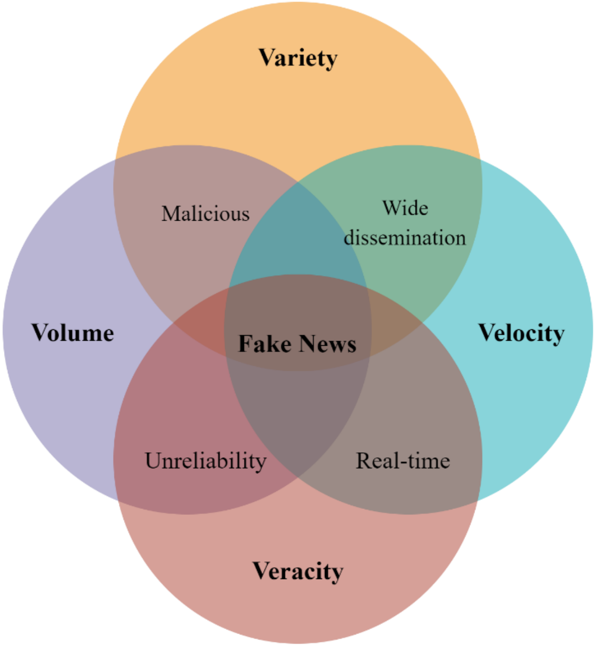

Figure 2. Different aspects of FN.

1.2. Psychology of FN

Users fall prey to FN easily. The main reason is cognitive biases (human thinking) that lead them to draw fallacious conclusions. News consumers may interpret information differently due to established cognitive biases [54]. Four types of cognitive biases are particularly pertinent in connection to FN. Primarily, users scan headlines without reading the related content [55], which explains the popularity of clickbait depending on attention-grabbing headlines. It proves to be detrimental to the dissemination and impact of false information. The users also comprehend and focus on shared news according to indications of popularity (bandwagon effect) on OSNs. Thus, popularity bypasses news legitimacy [56]. Third, partisanship is a strong response that bogus news exploits. The established beliefs and associations prevent users from fathoming the actual information [57]. In addition, erroneous information has a peculiar propensity to endure even after it has been corrected, i.e., belief echoes, which brings the users to the final point [58]. The users exchange online news in anticipation of social interaction, self-expression and status-seeking. The research demonstrates that people distribute fake information on social media automatically and reactively rather than intentionally. Occasionally, false information is spread deliberately to inform, enlighten and warn others [59].

1.3. Economics of FN

The emergence of digital technologies has changed the news production and distribution pattern from a linear economic model to a multisided algorithm-driven model [60]. According to Reuters Digital Report, two-thirds of web users access news via search engines, OSNs, or other online sources [60]. The statistics show that information access, communication, and reading news are three major pursuits of internet users. Additionally, the number of online news readers will rise to around 1.88 billion users worldwide by 2027, illustrated in Figure 3. Consequently, it will result in a high volume and frequency of FN [60].

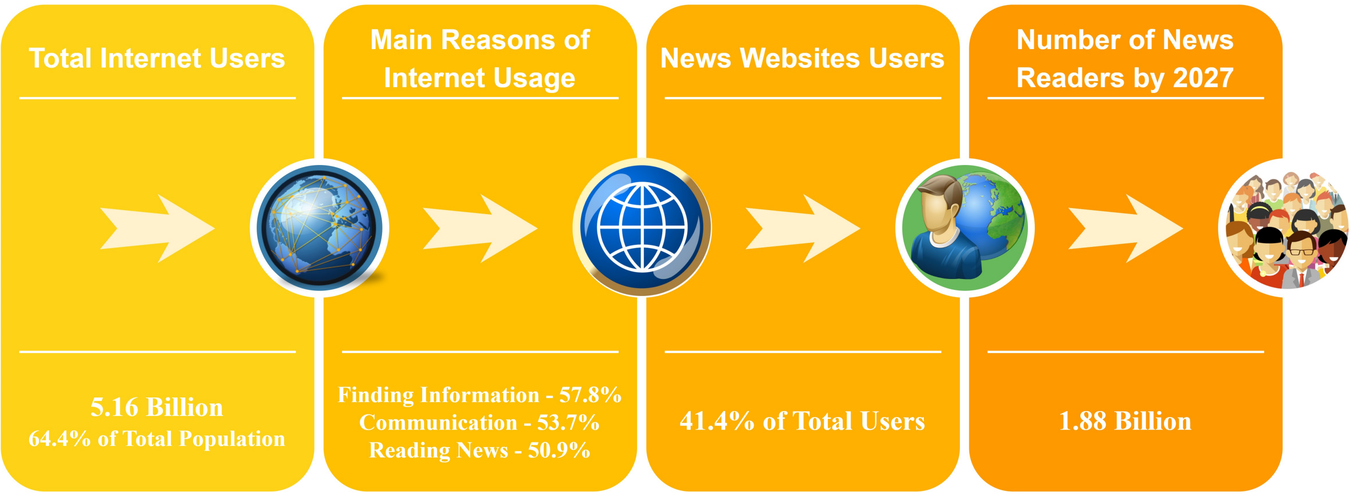

Figure 3. Statistical overview of digital news readers worldwide [61].

The algorithm-driven approach on OSNs and online platforms is used to attract users. The users are presented with FN, and they share FN and dubious information in their social circles. The platforms gain more advertising revenue through increased organic traffic [62]. Fake claims, clickbait and questionable content lead to trust issues in news from traditional media. According to Ipsos, the prevalence of FN on OSNs and online platforms have contributed to highly declined trust in traditional media over the previous five years [63]. Contrarily, FN production and distribution are far more simple and affordable. Moreover, FN creators hardly face legal repercussions [64]. Lastly, in addition to algorithm-driven platforms, cognitive biases like novelty and confirmation bias are crucial to FN consumption and dissemination, exploited by these algorithm-driven news markets [62]. Human cognitive bias, ineffective legal regulations, and financial incentives for algorithm-driven news curators to spread FN have all contributed to increased FN volume and velocity. Thus, this review expands on the previous literature in automatic FND using ML, DL, and NLP models to identify challenges and future research directions.

1.4. Motivation

FN dissemination has diverse effects on the cognitive ability of the users that prompt users to share FN inadvertently. The existing literature on FND is based on different feature perspectives and FND techniques. However, the FND process can be addressed by evaluating the existing interdisciplinary studies, datasets, and approaches for FND detection. The existing review on FND determined fundamental theories and detection approaches [12] and disregarded dataset evaluation for various SOTA approaches as shown in Table 1. The review classified FND techniques into four categories, i.e., style, knowledge, propagation, and credibility [65]. The author of [66] categorised FND techniques further into social context and news content. The literature enhances the research and offers insights. However, with the emergence and evolution of AI-based methods, it is significant to summarise the SOTA techniques for FND. The users remain the chief source of sharing FN on OSNs rather than bots and trolls. Therefore, it is necessary to identify the motives that prompt them to share FN. In this review, firstly, we provide insight into the stimulus behind intentional and unintentional FN sharing on online platforms and OSNs. Different interdisciplinary studies explain these motives. Secondly, the linchpin of the FND process is the availability of large-sized, labelled, and publicly available datasets, which are assessed. Lastly, various SOTA approaches highlight various feature information in different ways depending on the type of learning. Thus, we review and describe the present status of SOTA approaches for FND in this review with a new perspective. The review paper flow is demonstrated in Figure 4.

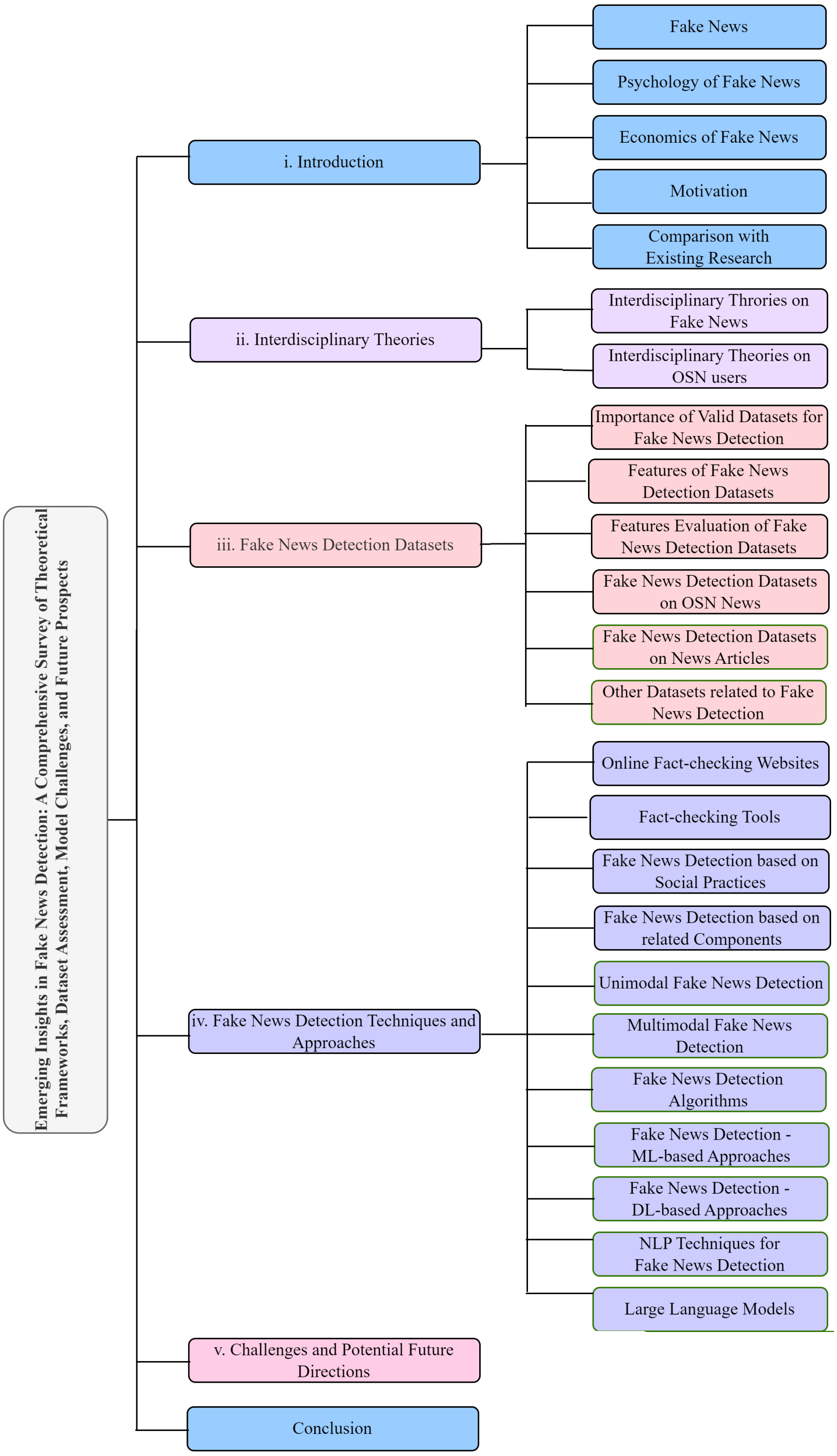
Emerging Insights in Fake News Detection: A Comprehensive Survey of Theoretical Frameworks, Dataset Assessment, Model Challenges, and Future Prospects
i. Introduction
Fake News
Psychology of Fake News
Economics of Fake News
Motivation
Comparison with Existing Research
ii. Interdisciplinary Theories
Interdisciplinary Throries on Fake News
Interdisciplinary Theories on OSN users
iii. Fake News Detection Datasets
Importance of Valid Datasets for Fake News Detection
Features of Fake News Detection Datasets
Features Evaluation of Fake News Detection Datasets
Fake News Detection Datasets on OSN News
Fake News Detection Datasets on News Articles
Other Datasets related to Fake News Detection
iv. Fake News Detection Techniques and Approaches
Online Fact-checking Websites
Fact-checking Tools
Fake News Detection based on Social Practices
Fake News Detection based on related Components
Unimodal Fake News Detection
Multimodal Fake News Detection
Fake News Detection Algorithms
Fake News Detection - ML-based Approaches
Fake News Detection - DL-based Approaches
NLP Techniques for Fake News Detection
Large Language Models
v. Challenges and Potential Future Directions
Conclusion

Figure 4. Review paper flow.

Table 1. Comparison with the existing reviews on Fake News Detection.

Existing Reviews	Year	Fake News Typology	Interdisciplinary Theories	Dataset Evaluation	FND Expert-Based Approaches	FND Feature-Based Approaches	Unimodal and Multimodal Approaches	Fake News Detection Techniques				
								ML	DL	NLP	Pre-Trained Transformers	LLMs
[67]	2018	✗	✗	✓	✗	✓	✗	✓	✗	✓	✗	✗
[65]	2018	✗	✓	✗	✓	✓	✗	✗	✗	✗	✗	✗
[68]	2019	✓	✗	✓	✗	✗	✗	✓	✓	✓	✗	✗
[69]	2019	✓	✗	✓	✗	✗	✗	✓	✗	✗	✗	✗
[48]	2020	✓	✗	✓	✗	✓	✗	✓	✓	✗	✗	✗
[70]	2020	✓	✗	✓	✓	✓	✗	✗	✓	✗	✗	✗
[71]	2021	✓	✗	✓	✗	✓	✗	✗	✓	✗	✗	✗
[72]	2021	✓	✓	✗	✗	✓	✗	✗	✗	✓	✗	✗
[73]	2021	✓	✗	✗	✗	✓	✗	✗	✓	✗	✗	✗
[74]	2021	✓	✗	✓	✓	✓	✗	✗	✓	✗	✗	✗
[75]	2021	✗	✗	✓	✗	✓	✗	✗	✗	✗	✗	✗
[76]	2021	✓	✗	✓	✓	✓	✗	✗	✗	✗	✗	✗
[77]	2021	✓	✗	✗	✗	✗	✗	✓	✗	✓	✗	✗
[78]	2021	✓	✗	✓	✓	✓	✗	✗	✗	✗	✗	✗
[79]	2022	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗
[80]	2022	✓	✗	✓	✗	✓	✓	✗	✗	✓	✗	✗
[81]	2022	✓	✗	✓	✓	✗	✓	✓	✗	✓	✗	✗
[82]	2023	✓	✗	✗	✓	✓	✗	✗	✓	✗	✗	✗
[83]	2023	✓	✓	✗	✓	✓	✓	✓	✓	✓	✗	✗
[84]	2023	✓	✗	✓	✓	✓	✗	✗	✓	✓	✗	✗
This review		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

2. Research Methodology

The literature review aimed to collect, identify, and report relevant research studies on FND in a structured and transparent manner, as explained in Figure 5. We referred to [85] and employed a structured procedure to identify, include, and exclude the relevant studies for this review and address the issue of automated FND. The research studies were identified using the keywords "Fake News Detection", "Interdisciplinary Theories on Fake News and users", "Fake News Detection using Machine Learning", "Fake News Detection Datasets", etc. Multiple search engine publishers' databases, such as IEEE, Elsevier, Springer, etc., were used for this process. The identified 450 research studies included early access publications, conference papers, and journal articles. The editorial papers were excluded after reviewing the abstracts, titles, and keywords. Therefore, we selected 355 relevant studies, which included conference proceedings, journal articles, etc. The number of these research studies is shown in Figure 5, respectively. The selected studies were used to evaluate the previous works in the literature to identify the research gaps and present future research directions that may assist in the automated FND process.

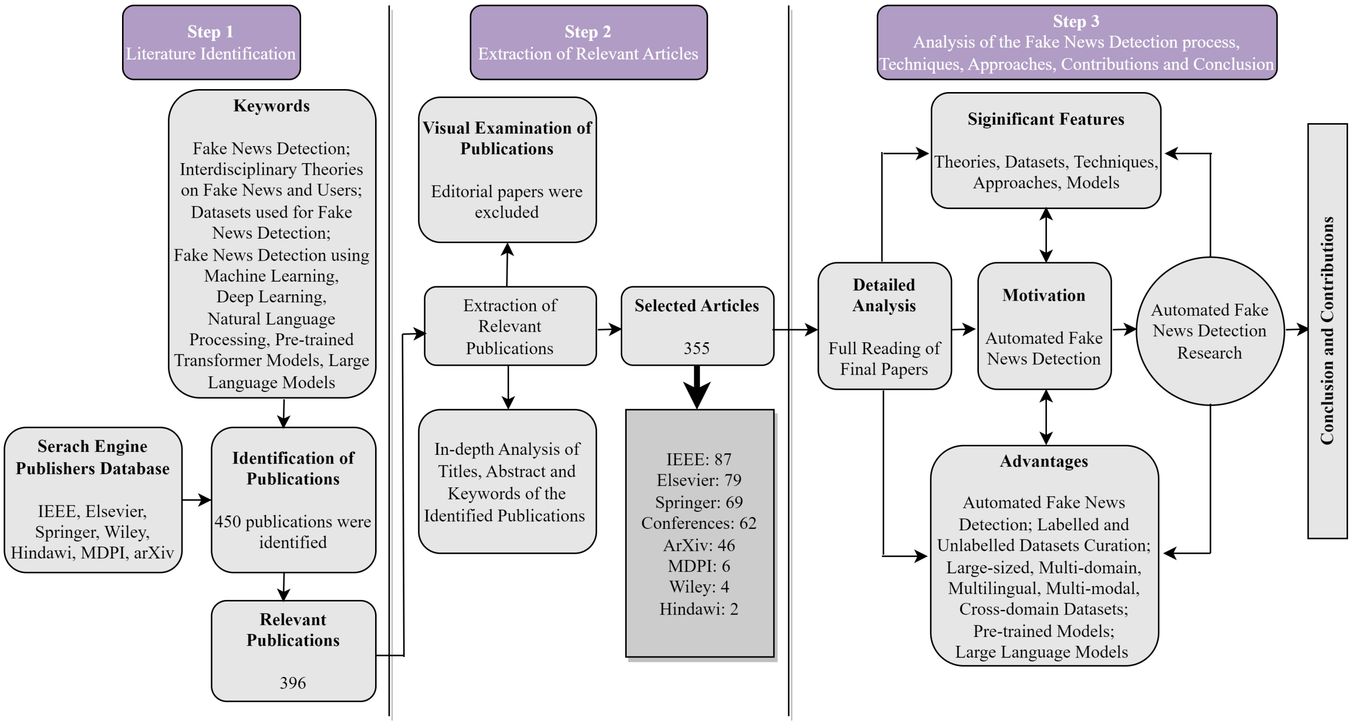
Step 1
Literature Identification
Keywords
Fake News Detection; Interdisciplinary Theories on Fake News and Users; Datasets used for Fake News Detection; Fake News Detection using Machine Learning, Deep Learning, Natural Language Processing, Pre-trained Transformer Models, Large Language Models
Serach Engine Publishers Database
IEEE, Elsevier, Springer, Wiley, Hindawi, MDPI, arXiv
Identification of Publications
450 publications were identified
Relevant Publications
396
Step 2
Extraction of Relevant Articles
Visual Examination of Publications
Editorial papers were excluded
Extraction of Relevant Publications
In-depth Analysis of Titles, Abstract and Keywords of the Identified Publications
Selected Articles
355
IEEE: 87
Elsevier: 79
Springer: 69
Conferences: 62
ArXiv: 46
MDPI: 6
Wiley: 4
Hindawi: 2
Step 3
Analysis of the Fake News Detection process, Techniques, Approaches, Contributions and Conclusion
Siginificant Features
Theories, Datasets, Techniques, Approaches, Models
Detailed Analysis
Full Reading of Final Papers
Motivation
Automated Fake News Detection
Automated Fake News Detection Research
Advantages
Automated Fake News Detection; Labelled and Unlabelled Datasets Curation; Large-sized, Multi-domain, Multilingual, Multi-modal, Cross-domain Datasets; Pre-trained Models; Large Language Models
Conclusion and Contributions

Figure 5. The methodology of the literature review and research process.

The review will determine FND and mitigation with the three most significant aspects. The overall contributions of the literature are summarised as follows:

- We evaluate the interdisciplinary theories on FN and OSN users. An in-depth assessment of various interdisciplinary theories on FN and OSN users explains the effective and explainable efforts for FND. The analysis of the datasets, research techniques, and technological models currently used in FND research is presented. The application of multimodal technology creatively describes and evaluates the advancement of scientific inquiry in communication, linguistics, and other FND-related fields.
- The online fact-checking sources and their limitations are reviewed, which lead to developing general techniques for automated FND categorised based on the stages of development. In addition, it discusses the research on the explainable model structure and behaviour, and evaluates explainable FND.

- The identified limitations, challenges, research gaps, and future research directions for FND using multilingual, cross-domain, and cross-lingual datasets, and SOTA techniques are presented. It highlights the prime methods and techniques that can be implemented to generate effective FND and mitigation models for OSNs. Lastly, recommendations to address the existing limitations of FND-based research are provided.

The remainder of this review is organised as follows. We first retrospectively consider the interdisciplinary theories on FN and OSN users in Section 3. We present and summarise the existing datasets in Section 4. We summarise FND techniques and approaches in Section 5. We report the identify challenges and suggest potential future research directions for FND in Section 7. We finally conclude the review aligned with prospective research directions in Section 8.

3. Interdisciplinary Theories

FN is aimed at users' vulnerabilities and social connections. The users' intentional and unintentional disseminations of FN depend upon different factors. The peripheral difference between their intentions is studied, i.e., based on their social identity and knowledge about certain news [86,87]. Therefore, the importance of in-depth research on the factors that provoke them to share FN and behavioural changes cannot be denied [14]. FND studies are worthy of note, with the use of existing interdisciplinary behavioural studies proposed by the social sciences (psychology, forensic psychology, and philosophy) and economics that provide insights into human behaviours and tendencies. The suggested studies [88] show that datasets are analysed (qualitatively and quantitatively), and SOTA models are created by using the cognition and behaviour of the users. This aspect of FND analysis has been an under-researched area to date [89]. The literature review of these interdisciplinary studies presents that these theories are hallmarks in FND and its analysis. These theories provide a proactive approach (early FND) by detecting and studying its spreader's behaviour. The assorted table shows information about these theories about users, their social identity, and existing knowledge about the news and its content. The assorted Tables 2 and 3 show information about these theories regarding users, their social identity, and existing knowledge about the news and its content.

Table 2. Interdisciplinary theories on fake news.

Theories	Explanation	Reference
Phenomenon of fake news	Fake news has diverse effects on society, politics and economy. A different strategy could be adopted to combat fake news in each domain.	[90]
Undeutsch hypothesis	In terms of content style and quality, a statement based on genuine experience is different from purely speculative assumptions.	[91]
A rhetoric of motives	Information contrasts with communication content that causes emotional polarisation.	[92]
Reality Monitoring	Actual events are characterised by higher levels of sensory-perceptual information.	[93]
Erfurth's treatise on Surprise	Deception offers a way to imbalance an adversary through doubt.	[94]
Deception operations	Deception through a convincing story to support pre-existing beliefs.	[95]
Four-factor theory	The biased opinions and views in the FN are divergent and tilted towards a set objective.	[96]
Wittgenstein Philosophy	The sensory data that give us the current status on local conditions.	[97]
Information manipulation theory	Deception frequently contains great information and the biased opinions are divergent and tilted towards a set objective.	[98]
Source Credibility	The core of political communications is persuasion.	[99]

Table 3. Interdisciplinary theories on Online Social Network users.

Theories	Explanation	Reference
Bandwagon effect	People often act in a particular manner because follow the actions of others.	[19]
Normative and informational social influences theory	Users still make more mistakes than they believe their judgements to be even when normative social influence in favour of an incorrect judgment is entirely removed (as in the anonymous condition).	[100]
Social identity theory	Self-concept is based on one's perception of their affiliation with a certain social group.	[87]
Inoculation theory	The users change their behaviour towards new information if it contradicts their pre-existing knowledge, ethics and code of belief.	[101]
Internal and External factors	The widespread dissemination of fake information to influence perceptions and opinions is a result of both internal and external variables.	[72]
Availability cascade	People frequently adopt the views of others which are prevalent in their social groups.	[102]
Cognitive bias	Cognitive biases and internal controls directly affect the users' practices.	[103]
Confirmation bias	People frequently believe information that supports their previous views or theories.	[14]
Conservatism bias	An inclination to continuing believing the existent information even when exposed to new information.	[104]
Echo chamber effect	Communication and repetition within a closed system can increase or strengthen existing beliefs.	[105]
Desirability bias	Users tend to consume information that appeals to them.	[106]
Semmelweis reflex	People frequently reject new facts because they conflict with traditional norms and beliefs.	[107]
Naïve realism	The ability to directly perceive objects as they actually are is made possible by our senses.	[108]
Attentional Bias	The reiterated concepts affect the perception.	[109]
Interference theory	Individuals make judgments based on their existing encounters.	[110]
Theory of cognitive dissonance	The users' beliefs, feelings and perceptions about relevance are inconsistent. Users who are experiencing cognitive dissonance experience stress.	[111]
Prospect theory	The use of biased assessments in decision-making.	[112]
Selective exposure	Users select and consume frequently shared information.	[15]
Validity effect	Following repeated exposures, people are more likely to assume that information is accurate.	[17]
Overconfidence effect	The users trust their comprehension, expertise and judgments.	[113]
Naïve Realism	People reject new facts because they rely on their existing knowledge and opinions.	[108]
Valence effect	People frequently overestimate the chance of positive outcomes vs. negative outcomes.	[114]
Illusion of asymmetric insight	People believe that they are more knowledgeable than other people.	[115]
Contrast effect	The improvement or impairment of cognition is caused by repeated or simultaneous exposure to a stimulus with a different value in the same dimension.	[116]

3.1. Interdisciplinary Theories on FN

The news is the key element in the FND. The user behaviour, interest, and news consumption are correlated [117]. FN affects the users; in response, the users share the news unintentionally. Hyper-partisans and pernicious users [118] create and spread hoaxes to instigate the emotions and attention of the users. They create sensationalism and anarchy in society for their benefit and monetary purposes. The content and context of such news are cunningly designed to cause unrest. The research in [118] has been conducted to reveal the truth about FN supported by interdisciplinary theories and that the explained actual events are characterised by higher levels of sensory–perceptual information. The theory explains the effect of FN on society, politics, and the economy. It further emphasises that due to the wide dissemination and acceptance of FN in different areas, a different strategy could be adopted to address the issue [90]. Information also has a third quality that is descriptive rather than hortatory or moralising, and it is disagreed that the reader must react to polarised information.Information contrasts with communication content that causes emotional polarisation [92]. The sensory data give us the current status of local conditions [97]. The core of political communications is persuasion [99]. The author explained that all major military operations succeed due to the element of surprise, dependent on secrecy and quick action [94]. Therefore, FN unbalances an adversary with the element of surprise. The author briefed that the target audience is deceived by using adequate effort and time to create a convincing story to support pre-existing beliefs. One of the most fundamental human tendencies that deception operations take advantage of is sensitivity to conditioning [95]. The established theories distinguish between fake and genuine news through their characteristics. These characteristics help to differentiate the original content and forged fabrication. The feigned writing style, sources, quality, comments, and facts reveal the FN [91]. The frequency of words, syntactic features, etc., are also used to detect FN. The biased opinions and views in the FN are divergent and tilted towards a set objective [96]. Information manipulation theory suggests FN identification through word counts [98]. Likewise, various theories on disinformation and FN also determine that FND is possible through its features.

3.2. Interdisciplinary Theories on OSN Users

The theories about user behaviour and intention towards certain news indicate their attachment and social context. Under the shadow of this behaviour and connection (social, political, religious, interests, etc.), the users share and like FN deliberately or involuntarily. Therefore, the content and context of news and the user exposure and participation are responsible for news consumption. They also share FN in their circle and seek comments from other users. Therefore, it is normal for users to become attracted towards certain news [119]. Inadvertently, the users also disseminate FN if they find it beneficial for other users [120]. The other broadcasters are intentional (bots and trolls) and used widely to execute malicious activities [121]. They conduct these activities for ideological and monetary purposes, to create unrest, and for ill-based benefits [122]. Trolls provoke, stimulate, and cause damage by sharing sensitive news (text, images, videos, etc.). FN is crafted for targeted users [123] from different social backgrounds, ethnicities and religious beliefs on OSNs.

The behavioural studies of normal users also indicate that users are not always motivated to spread FN. Fundamental theories about users' behaviour [19,87,100–102] suggest that multiple social factors affect their choices and decisions. FN consumption practices are the key role players, where users prefer and share news about their pre-existing knowledge, beliefs, etc. The other major factor is peer pressure. Meanwhile, both external and internal factors [72] are involved in the wide circulation of FN. The study shows that cognitive biases, confirmation bias [17,103], and internal controls directly affect the users' FN consumption practices. Inoculation theory [101] reflects that OSN users change their behaviour towards new information and news if they contradict their pre-existing knowledge, ethics, and code of belief. The algorithms in OSN create an echo chamber [105] for the users, record their preferences, and show news and posts accordingly [124]. The state of user refutation occurs when they experience and encounter unusual news that negates and conflicts with their pre-existing beliefs and attachments. They react to it through harsh comments that result in group polarisation [125]. The theory of cognitive dissonance [111] refers to the inconsistency between the beliefs, emotions, and perceptions of OSN users concerning relevance. In the case of cognitive dissonance, the users experience psychological tension. The users, thus, accept or deny FN [126] according to their beliefs to relieve and reduce stress and apprehension. Therefore, the acceptance of biased information and FN consumption are directly proportional and lead to desirability bias [106]. The users tend to release dissonance and indifferences through their selective choices and judgments [127]. Similarly, the interference theory [110] states that OSN users form impressions based on their early experiences and vice versa. The users decide on their FN consumption, and practice and promote FN intentionally or unintentionally [128]. It leads to a judgmental and biased process towards the news, irrelevant to the facts and legitimacy. The prospect theory [112] further extends the decision-making based on biased judgments and selective exposure [129] pertinent to the frequency and probability of FN consumption. The validity effect also demonstrates that frequent dispersal and exposure to FN incites OSN users to share it within their circles [17]. The users have greater confidence [108,113] in their understanding, knowledge, and judgments [107] and often refute the idea of new learning [104,130]. The users prefer news and information that keep them in an echo chamber [106]. Users' perceptions of FN are based on events [109]. The users share FN as they find it trending [19] and often overlook its authenticity and legitimacy.

4. FND Datasets

The FND process depends upon two different features (input types). Ref. [66] determined these features to be news content and social context features. Therefore, the news items in a dataset refers to the data, articles, materials, etc., and additional information related to social background, social affiliations, etc. However, the availability of large-sized, multidomain, and labelled datasets makes automated FND challenging. It remains a substantial impediment for low-resource languages, as there is a lack of available online sources. The dataset curation is an upheaval task. Data are extracted from various online resources and organised as a dataset. The dataset curation process is shown in Figure 6. The public datasets are categorised into data from OSNs, articles, and claims. Compared to news articles, the short claims are one or a few sentences long and provide valuable information. The fake claims function as clickbait to attract users' attention and are far more effective than news articles [131]. ONS news items are longer than claims. These also contain non-textual data and structured data from accounts and posts. There are various online sources for data scraping, such as OSNs (Twitter, Facebook, WeChat, etc.), online news sources (BBC, Guardian, Times of India, BBC Urdu, etc.), fake news sources (InfoWars, Before its news, Ending the News, vishvasnews.com, etc.), satire sources (Faking news, The Onion, Satire Wire, Beaverton, etc.), and fact-checking sources (Politifact, Snopes, AFP, Geo News Fact check, etc.). The privacy restrictions on available online sources further obfuscate the data collection process. Therefore, researchers are bound to buy data from the available sources or through crowdsourcing websites. Some publicly available datasets are not fact-checked and heavily depend on the claims with crowdsourced veracity labels. As a result, biased data are generated, i.e., data restricted in diversity, volume, and quality. The existing datasets are used for the selected research directions and requirements. The quality of the annotations is assessed using evaluation metrics. In this review, we present and compare 59 existing datasets based on the input items, dataset size, language, labels, and annotation. The comparison analysis is presented in Table 4.

Table 4. Fake news datasets (multidomain, multimodal, and multilingual).

Datasets	Year	Sample Size	Domains	Content Type	Platform	Language	Labels	Annotators	References
FakeNewsNet	2018	422	Politics, Society	Fake News Articles	Mainstream Media, Twitter	English	2	Experts	[132]
Fakeddit	2020	1,063,106	-	Posts	Reddit	English	2, 3, 6	Experts	[133]
Twitter-15	2017	1478	-	Posts	Twitter	English	4	Twitter	[134]
Twitter-16	2017	818	-	Posts	Twitter	English	4	Twitter	[134]
Reddit_comments	2020	12,597	-	Posts	Reddit	English	2	Emergent, Politifact, Snopes	[135]
PHEME	2016	330	Society, Politics	Posts	Twitter	English	3	Crowdsourcing	[136]
PHEME-update	2018	6425	Politics, Society	Threads	Twitter	English	3	PHEME	[137]
FACTOID	2022	3.4 Million	-	Posts	Reddit	English	3	mediabiasfactcheck.com	[138]
MediaEval	2015	15,629	-	Posts	Facebook, Twitter, Blog Post	English	2	-	[139]
RUMDECT	2016	5442	-	Posts and Rumours	Twitter, Weibo	English	2	Twitter, Weibo	[140]
Rumor-Anomaly	2019	4 Million Tweets and 1022 Rumours	Politics, Science, Crimes, etc.	Threads	Twitter	English	6	Snopes	[141]
RumorEval2017	2017	297	-	Threads	Twitter	English	3	PHEME	[142]
RumorEval2019	2019	446	Natural Calamities	Threads	Twitter, Reddit	English	3	Politifact, Snopes	[143]
ComLex	2018	5303 posts and 2,615,373 comments	-	Posts and Comments	OSNs	English	5	Twitter, Facebook, etc.	[144]
Some-like-it-hoax	2017	15,500	Science	Posts	Facebook	English	2	Experts	[145]
BuzzFace	2017	2263	Politics	Posts	Facebook	English	4	Buzzfeed	[146]
Brazil-India Elec	2020	844,000	Politics (Messages during Elections)	Text, images, audios and videos	WhatsApp Groups	English	2	Experts	[147]
WeFEND	2020	65,132	-	Text, images	Twitter and Weibo	English	2	Trusted users	[148]
Yelp	2019	18,912	Technology	Text	Mainstream	English	2	Experts and crowdsourcing	[149]
MULTI	2020	1 Million	COVID-19	Tweets	Twitter	67 languages	2	Twitter	[150]
CoAID	2020	4251	COVID-19	Threads	Twitter	English	2	Politifact, FactCheck.org, etc.	[151]
COVID-HeRA	2020	61,286	COVID-19	Posts	Twitter	English	5	Experts, CoAID	[152]
HealthStory	2020	1690	Health	Threads	Twitter	English	2	HealthNewsReview	[153]

Table 4. *Cont.*

Datasets	Year	Sample Size	Domains	Content Type	Platform	Language	Labels	Annotators	References
HealthRelease	2020	606	Health	Threads	Twitter	English	2	HealthNewsReview	[153]
COVID-19-rumor	2021	2705	COVID-19	Tweets, News	Twitter, Websites	English	2	Snopes, Boomlive, Politifact	[154]
COVID-19 FND	2021	10,700	COVID-19	Tweets	Twitter	English	2	Twitter	[155]
CHECKED	2021	2104	COVID-19	Threads	Weibo	English	2	Sina Community Management	[156]
MM-COVID	2020	11,173	COVID-19	Threads	Twitter	English, Hindi, Portuguese, Spanish, Italian, French	2	Snopes, Poynter	[157]
CONSTRAINT-2021	2021	10,700	COVID-19	Posts	Twitter	English	2	Snopes, Politifact	[158]
Indic-covid	2020	1438	COVID-19	Posts	Twitter	Hindi, Bangali	2	Experts	[159]
ArCOV19-Rumors	2020	162	COVID-19	Threads	Twitter	Arabic	2	Fatabyyano, Misbar	[160]
COVID-Alam	2021	722	COVID-19	Tweets	Twitter	English, Arabic	2	Experts	[161]
COVID-19-FAKES	2020	3,047,255	COVID-19	Posts	Twitter	English, Arabic	2	UN, UNICEF, WHO	[162]
Politifact Fact Check	2014	221	Society, Politics	Text	Websites	English	5	Channel 4, Politifact	[163]
TI-CNN	2018	20,015	-	Text, images	News Websites	English	2	-	[164]
LIAR	2017	12,836	-	Text	Politifact	English	6	Politifact	[165]
Breaking!	2019	679	2016 United States Election	Text	News Websites	English	3	BS Detector	[166]
GossipCop	2021	19,759	Entertainment	Text	Websites	English	2	GossipCop, E!Online	[167]
TSHP-17	2017	10,483	-	Text	Website	English	6	Politifact	[168]
Buzzfeed	2017	71	2016 United States Election	Text	Websites	English	2	Buzzfeed website	[169]
Gandhi’s dataset	2020	46,700	-	Text	News Websites	English	2	-	[170]
Burfoot Satire	2009	4233	Society, economy, politics, technology	Text	Mainstream Media	English	3	-	[171]
Kaggle_UTK	2018	25,104	-	Text	-	English	2	-	[53]
FakeNewsAMT	2017	480	Education, Sports, Entertainment, Politics, Business, Technology	Text	Websites	English	2	Crowdsourcing	[172]
Celebrity	2017	500	Celebrity	Text	Website	English	2	GossipCop	[172]

Table 4. *Cont.*

Datasets	Year	Sample Size	Domains	Content Type	Platform	Language	Labels	Annotators	References
Ahmed's dataset	2017	25,200	-	Text	Websites	English	2	Politifact	[173]
MisInfoText-Snopes	2019	312	-	Text	Website	English	5	Snopes	[174]
fauxtography	2019	1233	-	Text	Website	English	2	Snopes	[175]
NewsBag++	2020	389,000	-	Text, Images	Websites	English	2	Experts	[176]
Fake_or_real_news	2019								[177]
FA-KES	2019	804	Syrian War	Text	Print Media	English	2	Experts	[178]
Japanese Fake News dataset	2022	307	-	Text, Images	Fact Check Initiative Japan	Japanese	2	Experts	[179]
Spanish Fake News Corpus	2019	971	Science, sport, economy, education, entertainment, politics, health, security, society	Text	Leading news and fact-checking websites	Spanish	2	Journalist	[180]
Bend the truth	2020	900	5	Text	Leading Newspaper websites	Urdu	2	Experts	[181]
Dataset for Pakistani Political Discourse	2023	49 Million	Politics	Text	Twitter	Urdu, English	2	Twitter	[158]
FakeCovid	2020	5182	COVID-19	Text	92 Fact-checking websites	40 languages	2	Experts	[182]
FEVER	2018	185,445	Wikipedia	Claim, Wikipedia data	Wikipedia	English	3	Experts	[183]
FEVER 2.0	2019	1174	Wikipedia	Claim	Wikipedia	English	3	Experts	[184]
FEVEROUS	2021	87,062	Wikipedia	Claim	Wikipedia	English	3	Experts	[185]
Emergent	2016	300	-	News Articles, Claim	Websites	English	3	Hoaxalizer, Snopes	[186]
MultiFC	2019	36,534		Claim	Fact-checking websites	English	2-40	-	[187]
Snopes dataset	2017	4856	-	Text, Images	Snopes	English	3	Snopes	[188]
IFND	2021	56,868	Politics	Text, Images	Fact-checking websites	English	3	Fact-checking websites	[23]
FACTIFY	2022	50,000	-	Tweets, posts	Twitter	English	2	Twitter, Experts	[189]

Table 4. *Cont.*

Datasets	Year	Sample Size	Domains	Content Type	Platform	Language	Labels	Annotators	References
WELFake	2021	72,134	-	News Articles	Kaggle, Reuters, McIntire, BuzzFeed Political Datasets	English	2	Reuters, Kaggle Experts,	[190]
FakeSV	2023	3654	-	Audio, Video, Text	Douyin, Kuaishou	Chinese	2	Manual	[191]
Kishwar's Pakistani News	2023	11,990	Pakistan, Cities, personalities and politicians	News Articles	Fact-checking websites	English	2	No information provided	[192]
MINT	2022	20,278	Multidomain	News Articles	Portuguese mainstream and independent media	Portuguese	5	Crowdsourced Annotations	[193]
Urdu at FIRE	2021	1500	5	News Articles	Leading Urdu news websites	Urdu	2	Expert Journalists	[194]
FAKES	2019	804	Multidomain	News Articles	-	English	2		[178]
CREDBANK	2015	60 Million	-	Rumours	Twitter	English	5	1736 Turkers	[64]
DanFEVER	2021	6407 claims	Wikipedia	Claim, Wikipedia data	Wikipedia, Den Store Danske	Danish	3	Expert Annotators	[195]
ISOT	2018	44,898	World News, Politics	News Articles	Legitimate News websites	English	2	Experts	[196]
Divide-and-Conquer	2022	28,334	5	News	Twitter	English	2	Twitter	[197]
Hindi Fake and True Dataset	2022	2178	5	News Articles	BBC-Hindi, NDTV	Hindi	2	No information provided	[198]
IBFND	2023	20,053	12	News	News Websites	Bangla	2	No information provided	[199]

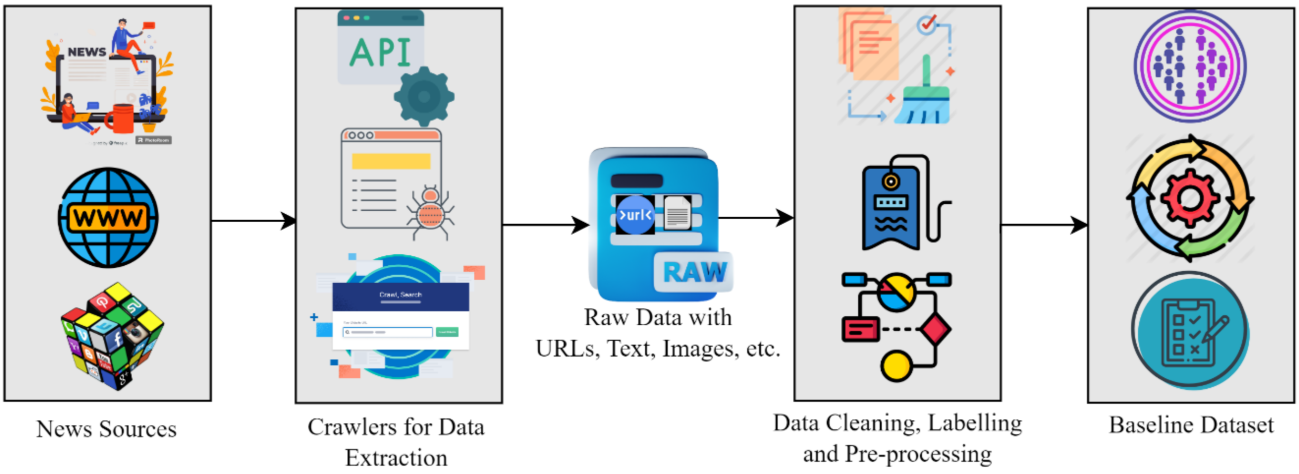
NEWS
WWW
API
Crawl, Search
>url<
RAW
Raw Data with URLs, Text, Images, etc.
News Sources
Crawlers for Data Extraction
Data Cleaning, Labelling and Pre-processing
Baseline Dataset

Figure 6. Dataset curation process.

4.1. Importance of Valid Datasets for FND

The menace of FN and its deep-rooted effects have gained popularity among researchers recently. FND and mitigation on OSN [200] are significant since FN may cause chaos and disturbance in society. On political grounds, if propaganda and negative perception (i.e., conspiracy theories) are maintained [14], it may result in information warfare between two states. Thus, the significance of FN to cause drastic effects on multiple levels and the infrastructure of a state cannot be negated [201]. The recent research aims to develop methods and techniques [69] to detect and mitigate FND. FND is possible using SOTA ML and DL techniques, where available datasets (labelled) are used to train the models and attain good results [202]. The main factor in FND is datasets. The dataset's validity and reliability are the foundations of FND and mitigation methods. The laborious work of FND is based on checking the accuracy and truthfulness of news and posts on OSNs. The big datasets on FN need to be accessed for early FND [69]. FND and determining the authenticity of FN in the shortest time possible are new research directions.

Thus, reliable dataset collection for FND is significant [69]. FN datasets are labelled based on the rating scale (veracity value), such as true, false, half-true, satire, etc. FN data are labelled manually, using assessment sites and computational tools (open web sources, knowledge graphs, etc.), or through crowd evaluation [33]. FN dataset labelling and analysis are faster and more convenient using digital tools as compared to manual annotation methods. However, ambiguous language and inconsistent annotations obstruct computational fact-checking and crowd evaluation. The four main categories of SOTA FND techniques are knowledge- and language-based approaches, topic-agnostic approaches, and hybrid approaches [202]. FND based on texts only is an Achilles' heel because of the complex nature of text-based posts and news. FND models categorise the news content and OSN posts as input data [203]. SOTA FND models analyse both the text content and context (user information [37], network information [204], publisher [205], etc.) of FN. Moreover, the legitimacy of newly created datasets and the creation of datasets is also significant. The review focuses on FND datasets for three different characteristics of FN and posts on OSN, i.e., style of the news, and user and network information. The chief source of FND research is the collection of reliable evaluation of labelled datasets as fake and genuine news. The collection of datasets is time-consuming. The developers construct datasets based on FN types, media, and topics. The developers select the characteristics and size of the datasets accordingly. FN extraction [69] is still an upheaval task because of the restrictions on OSNs and websites. FN types overlap due to content writing styles [33]. There is no set mechanism for FN classification. The review provides FN and related types of forged information to differentiate between FND models, techniques and results. FN datasets contain misinformation, propaganda, hoaxes, conspiracy theories, rumours, clickbait, and satire news [206]. The datasets may constitute different media, such as texts, images, and videos. Lastly, the selected topics include FN datasets on politics, finance, healthcare, etc.

4.2. Features of FND Datasets

- **News Content**: it comprises the language and syntax of the news content (headlines, stories and media) in the FND dataset.
- **Language**: the language used for the FN dataset as news is published in different languages.
- **Related Concept of Fake News**: the fabricated information and any form of FN along with FN articles, reviews on substandard goods, and commercials.
- **Specific News Categories**: FN datasets on different domains, such as healthcare, politics, economy, etc.
- **Media**: from conventional media to OSNs.
- **Availability**: free access to the dataset.
- **Application Purpose**: the main objectives of datasets can be FND, fact-checking, veracity value classification, and rumour detection.
- **Size**: the number of news items in a dataset.
- **Extraction Period**: period of news (news articles, images, comments, etc.) collection.
- **Rating Scale (Veracity Value)**: appropriate data labelling in the available dataset, such as false, mostly false, half true, mostly true, and true.
- **Integrity**: reliability (genuine or edited) of the dataset.

FND datasets have earned popularity due to their far-reaching effects on the users, society and infrastructure. The users are unable to exercise due diligence. Their cognition behaviour and lack of knowledge have further escalated the problem of FND on OSNs [69]. SOTA ML, DL, and NLP techniques are efficacious [69,200], where unedited FND datasets need to be leveraged. Therefore, the quality and legitimacy of datasets are the main building blocks. FND datasets from OSNs and online news articles are highlighted in the review.

4.3. Features Evaluation of FND Datasets

FN content is a specially designed type of hoax for illegitimate purposes in specific categories (political, financial, religious, etc.) to misguide OSN users. The distributed news and propaganda are attractive and tempting for the users. Different topics, media, and OSNs for intrusive FN are used [66], where datasets are the main ingredient of FND. FN for FND is classified based on multiple features (instigator and purveyor, content and context of FN, and audience). There are four salient [66,75] constituents of FN:

- **Instigator and purveyor**: The mastermind who creates FN content and broadcasts it to the target audience. The instigator and purveyor are malicious users with illicit intentions or bots, cyborgs, etc.
- **Content**: The news content includes concrete and abstract news features, i.e., text, media, and body of the news, along with intentions and objectives to create the content.
- **Context**: It represents conventional mainstream media, OSN platforms, and transmission strategies.
- **Target Audience**: OSN, traditional and mainstream media users.

The information about the instigator, purveyor, content and context of FN can be extracted from datasets. However, it does not apply to OSN users [89]. The main features of FN datasets are content and language, related concepts of FN (Table 1), specific news categories (politics, finance, religion, etc.), and a news rating scale (to evaluate genuine news). Additionally, the dataset size, availability, extraction period, and purpose of dataset application (e.g., FND, fact-checking, veracity classification, and rumour detection) [66], along with the type of online sources, are various constituents of FN datasets.

4.4. FND Datasets on OSN News

- **FakeNewsNet** [132] is the repository of updated datasets by Arizona State University for FND, misinformation, and rumours on OSNs [66,204]. Researchers have used the FND dataset for extensive research. The dataset contains multiple FN features, social context, geospatial details, etc. The rating scale of labelled data is two, i.e., false or true. The dataset is combined using fact-checked tweets and articles. Due to the restrictions in the privacy policy, Twitter API is used for data retrieval.
- **Fakeddit** is one of the largest publicly available multimodal datasets [133]. The dataset comprises various FN features, media, metadata and users' remarks on FN. The rating scale classifications are 2-way (true and false), 3-way (true, half true, and false), and 6-way (edited content, false association, fraudulent information, ambiguous information, satire, and true) with over a million threads.
- **Twitter-15 and Twitter-16** datasets comprise 1478 and 818 threads, respectively [8,134]. The veracity value of the datasets is false, half-true, true, and unverified. However, with the changes in Twitter's privacy policy, the researchers cannot use the datasets directly. Therefore, they may use an Application Programming Interface (API) to collect FND dataset subsets. The datasets are updated and constitute news-related tweets, reposts, replies, etc. The repository contains labelled Twitter propagation threads, highly retweeted messages and rumours based on veracity values.
- **Reddit_comments** [135,207] is an FND dataset comprising 2.86 billion records (709 GB in size).
- **PHEME** is used by researchers for FND on OSN (Twitter) to detect rumour-related posts [135]. It consists of 330 threads and rumours on nine different events. The related threads and rumours are annotated as false, true and unverified. PHEMEupdate is the updated version [137]. The number of total threads is 6045 threads, and the number of rumours is extended to 1067 true, 638 false, and 697 unverified rumours, respectively.
- **FACTOID** dataset is curated for political FN [138]. The primary source of annotated news is mediabiasfactcheck.com. It comprises 3.4 million Reddit posts from January 2020 to April 2021 by 4150 users. The dataset provides precise labels along with the users' political bias (extreme left to extreme left) and credibility level (low to high).
- **MediaEval** is a dataset used to retrieve, analyse and improve existing algorithms for multimedia access. The other datasets used in the previous research on FND suggest scope for improvement by adding new threads. The benchmarking initiative involves researchers on multimodal approaches to multimedia (sensor data, videos, images, etc.). In the competition, they use the MediaEval dataset [139,208]. The primary focus is multimedia and its effects on the users, society, and the multimedia system. The dataset contains posts from OSN (Twitter) labelled false and true.
- **RUMDECT** is a dataset that contains posts related to rumours and non-rumours [140]. The dataset is a collection of posts from Twitter (778 reported events) and Weibo (2313 rumours and 2351 non-rumours). Tweets are collected during the period of March–December 2015
- **Rumor-Anomaly** is a dataset that is a combination of the PHEME [136] and FakeNewsNet [66,204] datasets. The Snopes rating is used for thread annotation (stance and veracity) [141]. The dataset contains the tweets (4 million) of users (3 million), with articles (305,115) and hashtags (28,893) revolving around 1022 rumours (from 1 May to 1 November 2017).
- **Rumor-Anomaly** is an extension of the Rumor-Anomaly dataset, i.e., RumorEval2017 [142] and RumorEval2019 [143], and contains Reddit data (posts, users' information) as well. These datasets are used in the RumorEval workshop. Three labels (false, true and unverified) are used for veracity categorisation, and four labels (deny, comment, query, and support) are used for stance categorisation.
- The proposed dataset in the study [144] comprises a large set of comments (2,615,373) and OSN posts (5303). It is based on the linguistic analysis of users' comments on OSNs (Twitter, Facebook, etc.). The assigned annotations are (false, mostly false, false, mostly true, and true).

- Another dataset, Some-like-it-hoax, contains false news and posts on Facebook [145]. It comprises posts (15,500) from Facebook users (909,236) (from 1 July 2016 to 31 December 2016). The veracity value of the liked posts on Facebook is hoaxes and non-hoaxes.
- **BuzzFace** (based on the BuzzFeed datase) contains United States election news on Facebook from nine news agencies [146]. The dataset comprises text and media (images and movies) on fake and real news (2282 news articles). The journalists have checked the facts in the post and related news articles. The news is classified as mostly false, a combination of true and false, mostly true, and fake content. Figure 7 provides a visual representation of the key benchmark datasets commonly used in this research field, shedding light on the diversity and scale of the data sources employed for evaluation.
- The dataset is based on fake news on WhatsApp during the Brazilian (2018) and Indian (2019) elections [147]. The dataset contains disseminated images of misinformation during the elections. The datasets are organised in veracity value of misinformation and not-misinformation.
- **WeFEND** is a multimedia dataset collected from Twitter and Weibo [148,209] for FND. Tweets are gathered from the MediaEval benchmark [210,211] and constitute text, images, and social context. For Weibo content, true news is compiled from Xinhua News Agency, China. The Weibo FN dataset is from the official FN Weibo system for exposing and verifying news (May 2016–January 2016), which is verified by trusted users. The Twitter dataset contains 6026 true news, 7898 FN, and 514 images. The Weibo dataset comprises 4779 true news, 4749 FN, and 9528 images.
- The **Yelp** dataset is structured to categorise fake and trustworthy reviews [149]. It is based on reviews on OSN Yelp from metropolitan cities (Miami, Los Angeles, San Francisco, and New York) in the United States. The dataset entities are fact-checked by experts and crowd-sourcing to detect fraudulent reviews. The dataset contains around 19 k entities, categorised as user-centric and review-centric. The review-centric category focuses on text in the shared reviews, and the user-centric category focuses on user profiles, social collaborations, shared reviews, etc.
- The **MULTI** dataset also encompasses 9528 posts related to rumours and non-rumours from OSN Weibo [150]. The multimodal dataset contains both texts and media. The Weibo dataset is a collection of 9528 posts from Sina Weibo, one of the largest OSNs in China [210,212]. The researchers access fake threads, posts, and responses (in the Chinese language) through Weibo API. The legitimacy of posts is certified by the Sina Weibo management. The posts are classified as real or false. There are several datasets on FN related to health. At the dawn of the COVID-19 pandemic [150], panic and terror caused the users of OSNs to fall prey to the FN of OSNs. The users shared information attributed to the pandemic and vaccines to help others, while abundant FN surfaced [213,214]. Fake multimodal posts caused fear among the users and had a diverse impact on society.
- The **FakeSV** dataset contains 1827 fake and 1827 true videos from Chinese OSN applications [191]. The authors collected the videos from Douyin (Chinese TikTok) and Kuaishou from January 2019 to January 2022. The dataset includes user title, video, title, metadata and comments. The dataset is manually annotated.
- **Divide-and-Conquer** is scraped from Twitter and contains 28,334 news instances [197]. The multidomain news is categorised as true and fake news. There are several datasets on FN related to health. At the dawn of the COVID-19 pandemic [150], panic and terror caused the users of OSNs to fall prey to the FN of OSNs. The users shared information attributed to the pandemic and vaccines to help others while abundant FN surfaced [213,214]. Fake multimodal posts caused fear among the users and had a diverse impact on society.
- **CoAID** is the amalgamation of COVID-19-related FN on OSNs and websites [151]. It contains OSN posts (926), news (4251), and user engagements (296,000) about COVID-19. The FN period selected for FN is from 1 December 2019 to 1 September 2020.

- **COVID-HeRA** created as an extension of the CoAID dataset, includes news and posts related to COVID-19 [152]. For example, the story about drinking bleach to get rid of COVID-19 posed a potential risk to the general safety of the public. The posts and news are categorised as refutes, highly severe information, possibly severe misinformation, not severe, and real news.
- FakeHealth created the HealthStory and HealthRelease datasets with news, posts, social engagements, etc. [153]. The HealthStory dataset comprises news (1690), tweets (384,073), responses (27,601), and re-tweets (120,709). The HealthRelease dataset contains news (606), tweets (47,338), responses (1575), and retweets (16,959). Due to the Twitter privacy policy, information about social engagements and user networks can be accessed through API. The data are annotated as true and false.
- **COVID-19-rumor-dataset** contains news and tweets on rumours about COVID-19 [154]. The dataset is labelled as false, unverified, and true. It also comprises information on users' stances and sentiments along with the news (4129) and tweets (2705).
- The **COVID 19-fake-news-detection** dataset contains tweets, posts, and news from OSNs and various fact-checking websites [155]. The data are labelled as fake and real. The dataset comprises only texts in the English language.
- Several multimodal datasets have been curated that are in different languages. The CHECKED dataset [156] contains misinformation on COVID-19 in the Chinese language on Weibo. It includes verified microblogs along with reposts (1,868,175), comments (1,185,702), and likes (56,852,736) from December 2019 to August 2020.
- The **MM-COVID dataset** is curated with a focus on news in six different languages [157]. It also contains news (3981 fake and 7192 real)and posts of different media types.
- The **CONSTRAINT-2021 Hindi Shared Task dataset** [158] contains total posts (8200), hostile posts (3834) and non-hostile posts (4358). The data are collected from OSNs (WhatsApp, Twitter, etc.) and different fact-checking websites and OSNs. It comprises two-dimensional evaluations, i.e., coarse-grained (hostile vs. non-hostile posts) and fine-grained evaluation (hostile classes). The dataset is labelled with entities (non-hostile, defamation, offensive, hate speech, and fake news).
- The **Indic-covid** dataset contains tweets in Hindi and Bengali language to detect FN on OSNs [159].
- The **ArCOV-19 dataset** comprises tweets in the Arabic language from 27 January to 30 April 2020 on COVID-related news. It contains around 1 million tweets and the most liked and retweeted posts and news.
- **ArCOV19-Rumors dataset** [160] is an extended version of the ArCOV-19 dataset. The dataset has 162 verified claims and labelled tweets with the veracity value (fake or true) on claim-level verification. The other classification mode is tweet-level verification with values (re-share and response).
- **COVID-Alam dataset** is a multilingual repository [161] of COVID-19-related data comprising over a million tweets (1,101,349) by users (344,328) during January–December 2021. The dataset is annotated in detail regarding questions including the following: To what extent does the tweet appear to contain false information? Will the tweet's claim have an impact on or be of interest to the general public?
- **COVID-19-FAKES** [162] also comprises multilingual tweets (3,047,255) on COVID-19. The dataset is labelled with the veracity values, misleading and real.

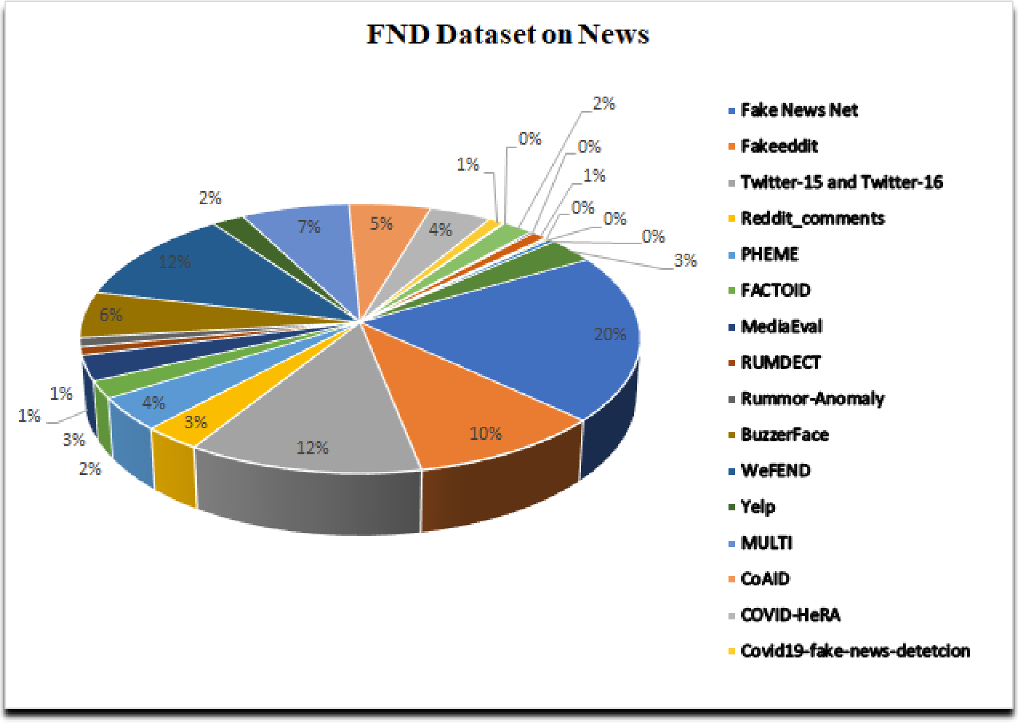

Figure 7. A pie chart of the benchmark datasets used in the studies of Fake News Detection.

4.5. FND Datasets on News Articles

- **Top-notch FND**: The Top-notch FND dataset, fact-checking from news articles, is based on the PolitiFact and Channel4 websites [69,163]. Experts have fact-checked 221 statements in the dataset. The dataset comprehends news reports on United States politics (i.e., lobbyists and congress affiliates). Along with text, the headlines are also part of the dataset, which provides insights into why users are attracted to such news. The dataset also includes text and fact-checking evidence from websites. Five-scale veracity divides FN data as false, mostly false, half true, mostly true, and true. The links of the FN sources and the fact-check article presented on Politifact have been published. The Politifact Fact Check dataset is also a corpus of 21,152 testimonials taken from the PolitiFact fact-checking website. The dataset is annotated by experts and grouped as pants on fire, false, mostly false, half true, mostly true, and true. Additional information as evidence sources is also included. A pie chart is shown in Figure 8, illustrating the benchmark datasets commonly employed in studies related to Fake News Detection.
- **TI-CNN** is the dataset of around 20 k news [164]. The news articles in the dataset are accumulated from various authentic news websites (Washington Post, New York Times, etc.). Expert journalists annotated the news articles. The news articles include the author, image or text, and title).
- The **Buzzfeed** News dataset is the corpus of news articles (1627) published during the 2016 United States election [169,215] on Facebook. The period of selected news is from 19–23 September to 26 and 27 September. Expert journalists annotated the dataset. The selected labels for the news articles are satire, real, and fake news.
- The **LIAR dataset** curates 12,836 claims on political news [165]. The assigned veracity values are pants on fire, false, mostly false, half true, mostly true, and true. The claims are collected from 2006 to 2017 and are manually labelled. The source of the statements is politifact.com. Experts have checked the statements. An extension of the LIAR dataset (LIAR-plus) used by [216] demonstrates that the users' justifications can be extracted for related claims.

- The **Breaking!** dataset is the consolidation of FN before and during the 2016 elections in the United States [166]. The linguistic features of FN are used for classification purposes. The news items are categorised as opinions, partial truth, and false.
- The **GossipCop** (Suggest) dataset is a collection of rumours and FN articles (19,759) on showbiz in the United States (July 2000–December 2018) [167]. FN in magazines and on websites are assembled in the dataset. The experts have organised the news tag annotations. The news and articles are rated on a 0–10 scale for the range of rumour to true news, respectively.
- The **TSHP-17** dataset is the fact-checking corpus for declarations of political leaders [168]. It comprises FN (33,063) related to propaganda, hoaxes, satire, and factual news articles. The news articles are taken from Politifact and other sources available for the news articles. The news items are categorised as pants on fire, false, mostly false, half true, mostly true, and true.
- The dataset contains fake and real news articles (46,700) [170]. The news is collected from Kaggle and leading news websites (Fox News, Washington Post, CNN, The Guardian, The Onion, BBC, etc.). It contains real news (22,056) and FN articles (24,194). The news for the dataset is compiled from 2015 to 2018.
- The **Burfoot** Satire News dataset is a repository of satire news related to society, technology, political and economic issues [70,169,171]. The dataset contains real news (4 k) and satire news (233) articles. The dataset has been amassed to distinguish between real news articles and satire. Satire is selected for each real news item, and real news is collected from the English Gigaword Corpus.
- **Kaggle–UTK**, a dataset on FN, contains 25,104 news articles [53]. The dataset for political FN, satire and clickbait has been created based on the list provided by [217]. It is a repository of websites and contains 225 k FN, real news, and satire in total. KDD 2021 TrueFact Workshop: Making a Credible Web for Tomorrow inaugurated the competition on Kaggle to develop models and datasets for FND [161].
- The **FakeNewsAMT** and Celebrity datasets were amassed by [172] for FND. For FakeNewsAMT, the news (240) was collected from leading news websites (CNN, The New York Times, ABC News, etc.). The news was amassed using crowdsourcing related to six different fields. The Celebrity dataset contains both true and fake news (500). Authentic news (250) is obtained from leading magazines (RadarOnline, People Magazine, etc.) and fake news (250) from GossipCop.
- The dataset with real news from world sources (25,200 news articles) comprises true and fake news articles [173]. Reuters News websites were used to collect real news articles, whereas FN was taken from the Kaggle dataset from unreliable websites fact-checked by Politifact.
- The **MisInfoText-Snopes** dataset includes 1692 verified news articles from Snopes [174]. The experts have annotated news articles and categorised them as fully false, mostly false, and a mixture of false, true, mostly true, and fully true. By contrast, MisInfoText-Buzzfeed contains 1431 news articles related to the United States election. The number of articles is segregated as mostly true (1090), mixture of false and true (170), mostly false (64), no evidence (56), and Buzzfeed news articles (33) [178].
- Two datasets are combined for fauxtography; as the name suggests, these are related to misleading images and claims [175]. The images and claims (1233) in the dataset were collected from Snopes and Reuters. There are 197 images with true claims from Snopes and 395 from Reuters. The images with fable claims (641) are taken from Snopes.
- The **NewsBag++** dataset is a repository of multimedia news articles [176]. The articles were scraped from The Onion and Wall Street Journal. It comprises both fake (389 k) and real news (200 k). The news is characterised as fake or true. The FA-KES dataset on news related to the Syrian war contains news from print media [178]. The labels were created by experts for 804 news articles. The veracity value of the dataset is selected as fake or credible.

- The Fake_or_real_news dataset contains news articles (6337) related to society and politics [177]. The ratio of fake and real news in the dataset is the same, which is gathered from mainstream media (no further information shared by the author). The news is categorised as fake or real news.
- The **Japanese Fake News dataset** is a consolidation of real news published in Fact Check Initiative Japan that aims to prevent society from the perilous effects of FN and spreading FN [179]. The news articles (307 posts) were congregated manually by Twitter search from July 2019 to October 2021. Through Twitter API, 186 FN and related contexts were also set up. The dataset includes 471,446 tweets from 277,106 users for 17,401 conversations.
- **Spanish Fake News Corpus** contains news articles (1233) in Spanish from leading news and fact-checking websites [180]. News articles were collected from November 2020 to March 2021 on nine various domains. The dataset comprises Spanish-v1 and Spanish-v2 with 971 and 572 news articles, respectively. The news articles are divided into fake and true news.
- The **"Bend the truth"** dataset contains 900 news on five different domains (politics, entertainment, etc.) in the Urdu language [181]. The news was collected from different Urdu news websites and manually annotated. The news is categorised as fake and real news. The fake news was written by journalists. A range of Urdu news websites (BBC Urdu, CNN Urdu, Dawn News, etc.) were used to extract news articles.
- The dataset for Pakistani Political Discourse [158] (after the No Confidence Vote) contains around 49 million tweets from 19 April to 7 May 2022. The multilingual data (tweets in Urdu and English) are a repository of the government change and users' reactions to the event. The data were collected from Twitter through API and comprise Urdu (34,588,431) and English (9,026,404) tweets from OSN users around the world.
- The multilingual dataset FakeCovid comprises 5182 news articles from 92 fact-checking websites [182]. The news articles are classified as fake and real. The CheckThat! dataset is based on news articles from the dataset, which are from 105 countries around the world.
- The *ISOT* dataset contains 45,000 news articles, and the ISOT lab at the University of Victoria curated the dataset [196]. The news covers politics and the world news domain from 2016 to 2017. The true news was collected from leading news websites, including Reuters. The fake news was collected from various sources flagged unreliable by PolitiFact.
- Several competitions and workshops are held on fact verification worldwide. The FEVER workshop inspires researchers to offer some cutting-edge datasets for collaborative projects. One of the largest repositories of claims (185,445) for fact-checking is FEVER [183]. The first FEVER shared task was created by editing phrases that were taken from Wikipedia, using previously processed Wikipedia data. The three-pronged approach is used for accuracy evaluation, i.e., NotEnoughInfo, refute, and support. The dataset is annotated with proofs and valid documentation.
- The 2019 shared task's breaker phase participants submitted 1174 claims for the FEVER 2.0 dataset [184]. The participants gathered information about the hostile situations that cause systematic errors. The dataset with 1174 claims was created by the participants. The veracity value was the same as in FEVER, i.e., (supported, refuted, and NotEnoughInfo). The novel correctly labelled claims that were not included in the original dataset were considered valid.
- **FEVEROUS**, with around 88 k valid claims, is also annotated with evidence and documentation [185]. Additionally, the information includes annotation metadata, e.g., time signatures, query terms, and page clicks.

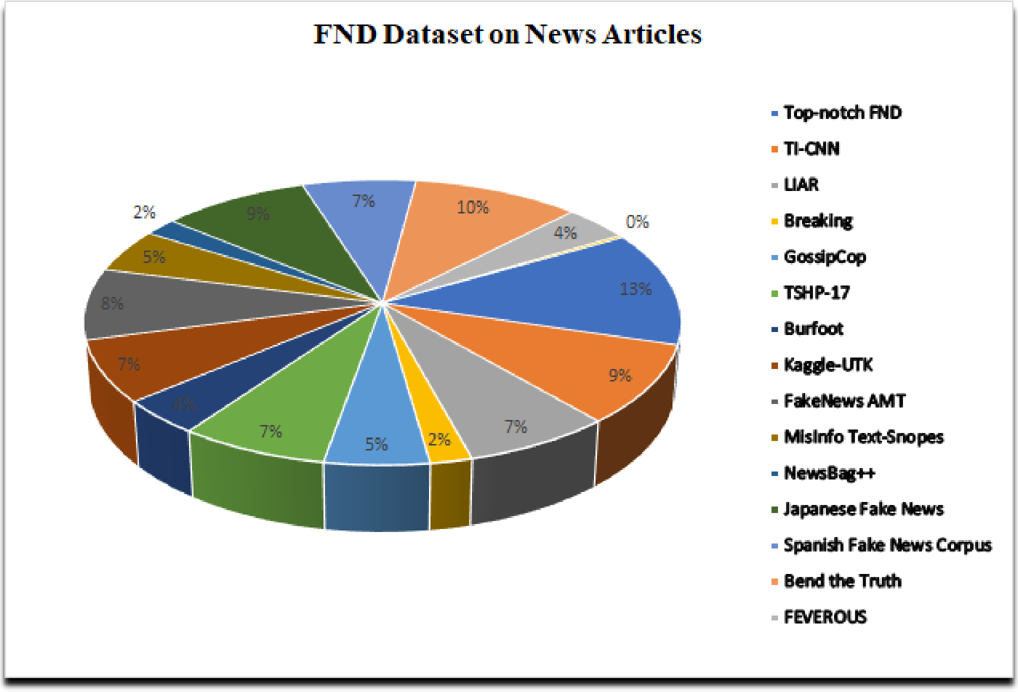

Figure 8. A pie chart of the benchmark datasets used in the studies of fake news articles.

4.6. Other Datasets Related to FND

- The **Emergent dataset** (FNC_dataset) comprehends 300 rumoured claims related to news articles (2595) [186]. The dataset is labelled with veracity values (true, unverified, and false). The dataset determines rumours and exposes the FN articles present on online media websites.
- The **MultiFC** dataset entails claims (34,918) accessed from 26 different fact-checking websites [187]. The details of the news along with rich metadata and veracity labels are included. The ranking scales are incorrect and correct and are used for the news. Along with the contemporary veracity values, 'grassroots movement!', 'misattributed', and 'not the whole story' are also used. Expert journalists have assigned labels to the news matter.
- The **Snopes dataset** covers 4956 claims [188] that are multimodal tweets and fact-checking articles. The dataset was collected from Snopes.com. The news items are labelled as unknown, false, and true. The UKP Snopes dataset is a repository of 6422 claims collected from the Snopes website [218]. The news in the dataset was gathered from news blogs and OSNs. The news is annotated as no stance, refutes, and agrees. The PolitiHop dataset [131] contains 500 fact-checking real-world claims with manual annotations.
- The **IFND dataset** is the amalgamation of Political FN and news on miscellaneous topics (56,868 news) in India [23]. The multimedia dataset (text, images, news headings, etc.) was collected from various fact-checking websites. The veracity values are real and fake news. The events in the dataset are from 2013 to 2021.
- **FACTIFY** is a multimodal fact verification dataset [189]. It includes textual claims (50 k), references, and multimedia (100 k). The dataset was formed to counteract FN in OSNs.
- The **Urdu at FIRE dataset** was presented by [194] and originally contained 900 fake and true news articles. The dataset was updated for shared tasks on Urdu FND, and the latest version consists of 1500 news articles from leading Urdu news websites. The news is from five different domains, and expert journalists annotated the dataset.
- The **CREDBANK dataset** contains around 60 million tweets from Twitter [64]. It is a repository of tweets related to events checked for credibility for set events (1049). Different annotators fact-checked and labelled the gathered events.

- The **Pakistani News** dataset contains 11,990 news articles in English from fact-checked news APIs [192]. The news covers politicians from Pakistan, of different provinces, cities, and personalities. The dataset is labelled as true and fake news.
- The **MINT** dataset comprises news articles in the Portuguese language [193]. The multidomain news was scraped from mainstream and independent media. The crowdsourced annotated dataset is labelled into five categories.
- The **Hindi Fake and True** dataset comprises 2178 news articles in the Hindi language from BBC-Hindi and NDTV [198]. The news in the dataset is assigned two labels.
- The **IBFND** dataset comprises 20,053 news in the Bangla language [199]. The fake and true news in the dataset are included from twelve different domains.
- The **DanFEVER** dataset is in the Danish language (6407 claims) [195]. Wikipedia data, Den Store, and Danske were scraped for the dataset curation. Expert annotators annotated the dataset, which is divided into three categories.
- **S**ome benchmark datasets were not given any name by their authors. Therefore, we have named such datasets after their first author or proposed model. Also, the pie chart in Figure 9 illustrates the benchmark datasets commonly employed in studies related to other datasets related to FND.

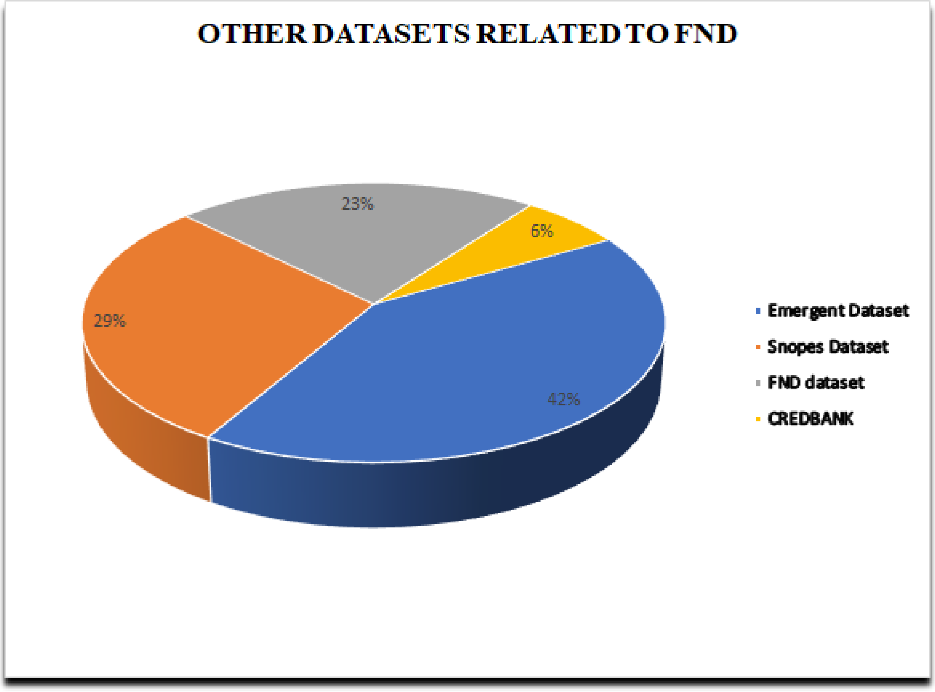

Figure 9. A pie chart of the benchmark datasets used in the studies of other datasets related to FND.

5. FND Techniques and Approaches

The existing research on FND is focused on investigating FN from a particular perspective, such as data mining and NLP. The previous literature classified FND techniques frequently depending on various ML-based approaches [74] and social contextual-based information [66]. In this review, we present the automated FND process based on the perspectives of OSN users (human-based techniques), SOTA AI-based techniques, and blockchain-based techniques. User-based techniques are fact-checking and crowdsourcing methods adopted by users to validate the news authenticity on OSNs and online websites. Various AI-based approaches have been proposed in previous works in the literature for automated FND. These include ML, DL, pre-trained transformer-based approaches, and NLP. Lastly, blockchain-based techniques establish the legitimacy of news sources and content transparency through blockchain methods. The FND techniques are further elaborated in the subsequent sections. An overview of the FND techniques and approaches is shown in Figure 10, and the FN features and FND approaches and techniques are presented in Figure 11.

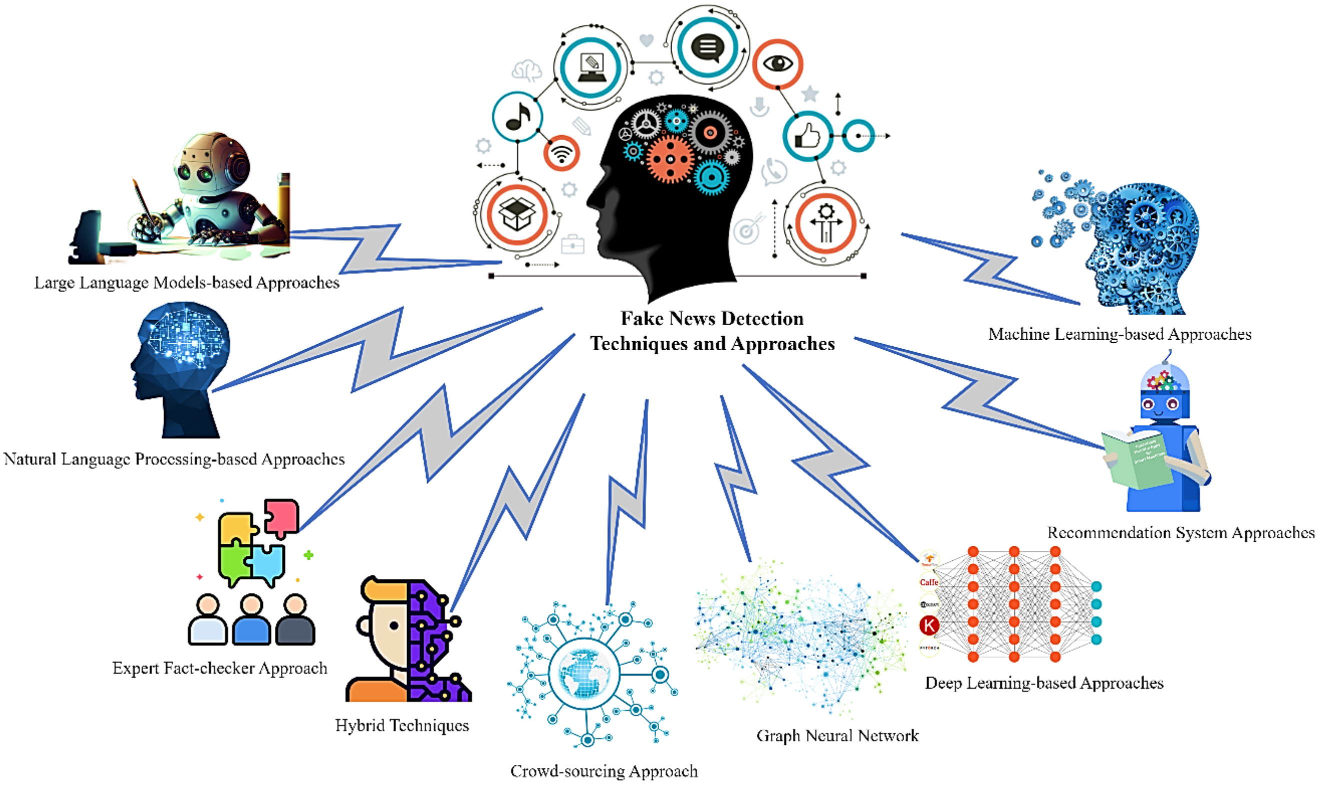
Fake News Detection Techniques and Approaches
Large Language Models-based Approaches
Natural Language Processing-based Approaches
Expert Fact-checker Approach
Hybrid Techniques
Crowd-sourcing Approach
Graph Neural Network
Deep Learning-based Approaches
Recommendation System Approaches
Machine Learning-based Approaches

Figure 10. Fake News Detection (FND) techniques and approaches.

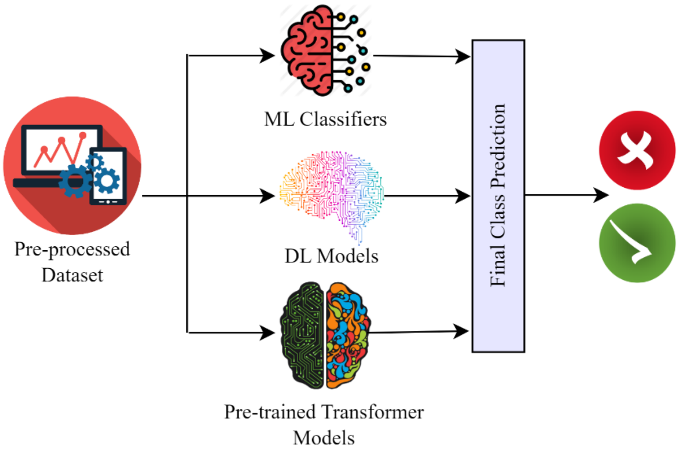

Figure 11. Different Fake News Detection approaches used for content-based analysis.

5.1. Online Fact-Checking Websites

Various third-party FND services are provided to users for fact-checking purposes. Due to the complexity of news items in terms of facts and figures, a single category outcome cannot evaluate and approve multiple news facts. The fact-checking services use different evaluations and graphical indicators to determine the accuracy of news. After due diligence, the fact-checking services label news items according to their veracity values, which helps readers distinguish between fake and true news. Different fact-checking websites included in the review are highlighted in Table 5.

Table 5. Online fact-checking websites: comparison and analysis.

Tools	Availability	Technique	Input	Output	Source
Fake News Detector	Browser Extension	ML, crowdsourcing	News content	Clickbait, extremely biased, fake news	Feedback produced by other tools
SurfSafe	Browser extension	Comparison and textual analysis	Text and images	Unsafe, warning, safe	Trusted organisations for fact-checking,
trusted news	Browser extension	-	Website content	Satire, biased, trustworthy	MetaCert protocol
FiB	Browser extension	Web scraping, text analysis, image analysis, semi-supervised ranking model	Posts from Facebook	Trust score	Image recognition, verification using keyword extraction, source verification
BS-Detector	Browser extension	Comparison model	URLs	Clickbait, conspiracy theory, extremely biased, fake news, etc.	Unreliable domains datasets
Fake News Guard	Browser Extension	Fact-checking, network analysis, AI	URLs, webpages	Fake or not	Fact checkers
BotOrNot	Website, REST API	Classification algorithm	Twitter screen name	Bot likelihood score	Accounts for recent history including tweet mentions
TrustyTweet	Browser extension	Media literacy	Tweets	Transparent and intuitive warnings, politically neutral	Potential indicators from previous studies
Decodex	Browser extension	-	Information fragments	Information, no information, satire	Online websites
LiT.RL News Verification	Web browser	Support Vector Machine, natural language processing	Language used	Clickbait, fake news, satirical news	Lexico-syntactic text features
TweetCred	Browser extension	Semi-supervised learning model	Tweets	Credibility score	Twitter dataset

- **FactCheck.org:** The non-profit online website authenticates the legitimacy of news and aims to reduce political misinformation and misunderstandings in the United States. Experts evaluate the accuracy of the claims and statements of prominent United States public figures. These claims and statements originate from media sources, including TV advertisements, speeches, in-person interviews, the most recent news, and OSNs. During election years, experts investigate the veracity of statements of party leaders. The experts and professionals ultimately validate the reliability of each shared claim. Additionally, the fact-checking service verifies scientific claims, health-related discussions and Facebook campaigns for FND.
- **PolitiFact.com:** The fact-checking website assesses the reliability of claims made by journalists, analysts, opinion writers, bloggers, and other public figures. The impartial online fact-checking website delivers election coverage. They classify the claims into seven categories. The experts evaluate the context and content of a statement before confirming the accuracy of the assertions and declarations. Thus, the users access verified claims, articles and improvements through the fact-checking interface.

- **Snopes.com:** Snopes.com is renowned and one of the first online fact-checking services for validating false statements in contemporary American culture. The fact-checking website covers different domains, including literature, economics, technology, crimes, scams, etc. Professionals verify the news. The website also offers fact-checked library collections by evaluations and veracity rankings. Leading newspapers in the United States and publications endorse their fact-checking services.
- **AFP Fact-check:** The fact-checking website analyses viral and influential online rumours and claims. The fake multidomain claims are circulated on blogs, OSNs, websites, and other public forums. They verify the claims from different regions worldwide. Their services are not limited to news and claims from a particular country or region.
- **Fact checker:** The Washington Post project validates the statements and facts in the speeches of leading politicians at local, national and international levels. The fact-checking forum aims to decipher the code words of the political figures and diplomats that mislead the readers and OSN users. Politicians use these obscure statements to hide the truth. The fact-checked news and claims are reported through extensive analysis and by adding missing context.
- **FactCheckHub:** The fact-checking service identifies and reports fake news, rumours, and claims. The non-partisan platform for fact-checked multidomain content provides verified news and claims.
- **Classify.news:** The online web source identifies FN claims and news articles. It aims to evaluate the news and posts' credibility based on ML techniques. It aggregates labelled news articles through examples from reliable websites. Two different categories of evaluation techniques, "Content-only" and "Context-only" are used. The former approach uses an iterative improving classification technique, and the second model is based on a multivariate regression Naïve Bayes.
- **OpenSecrets.org:** The politically neutral resource monitors the effect of used finances on biased governmental policies and United States elections. The web source provides updated information on fake, misleading, and perplexing news. FakeNewsWatch.com keeps track of hoaxes, fake reporting, and clickbait-based websites. Websites like fakespot.com and reviewmeta.com are well known for their web reviews and verification services. With the advancement of ML methodologies, Fakespot.com helps Amazon customers filter out fraudulent web comments, and reviewmeta.com uses analytical models to segregate fake customer reviews on products.
- **Hoax-Slayer.com:** The fact-checking service exposes online fraud and spam emails and enlightens users about general cybersecurity practices. They assist users in fighting against online criminal activities. Their services extend to the evaluation of online hoaxes based on reliable sources. The users may approach leading organisations, government agencies and relevant institutions to validate the legitimacy of particular communications.
- **Hoaxy.iuni.iu.edu:** The system gathers, analyses, and identifies misinformation and disinformation spread online and provides fact-checked analysis. An interactive visualisation demonstrates pre- and post-analysis of the publicly available tweets using fact-checked information. The system assists the users in identifying the authenticity of their preferred topics with unfounded claims and associated fact-checked data.
- **Factmata.com:** Factmata.com is a Google-funded program for claims detection and verification. The assertion verification project uses ML, AI methods and NLP techniques. The quantitative claims are verified through arithmetic relationship analysis. The website aims to expose bogus information by providing better factual and contextual information. Additionally, they can help businesses refrain from disseminating erroneous information and extremist material.
- **TruthOrFiction.com:** Online users can quickly and easily learn about e-rumours from this impartial web source. Their services also include the identification of emails with scams, virus warnings, and humorous or inspirational messages. The service

focuses on rumours circulated via emails and SMS. Additionally, they classify news and stories into seven categories.

- **Telugpost.com:** The online website verifies the misleading news and fake claims on OSNs and emails that may instigate violence and disturbance among the users and society. The unbiased selection and verification of memes, posts, claims, and viral images is ensured. The selected viral posts are multidomain. The experts monitor and evaluate the facts from authentic sources. Various reverse image search and video verification tools are used for multimodal claims and posts on the OSNs.
- **The Quint:** The internationally certified fact-checking service debunks multidomain misleading news and disinformation in Hindi and English. They follow a bipartisan approach to validate the credibility of multidomain news from India and the world.
- **Vishwasnews.net** The fact-checking website provides fact-checked news in 12 different languages including, Urdu, Hindi and English. The multidomain published news, clickbait and claims on media and OSNs are verified.
- **Soch Fact-check:** Fact-checked and verified news in regional, local, Urdu, and English languages can be accessed from the online web-source. They also debunk misleading statements and OSN posts of politicians and government officials. The project offers a limited number of news, claims, altered images, and videos. The news and claims are from different domains.

5.2. *Fact-Checking Tools*

The widespread use of OSNs and increased website content in the absence of checks and balances and lack of scrutiny serve as favourable conditions for vast FN dissemination. Thus, news legitimacy is greatly affected. The fact-checking services provide fact-checked and verified content. Different tools are also present that may clear the doubts of the users when faced with manipulated information, news and claims. The fact-checking tools included in the review are highlighted in Table 6.

Table 6. Online fact-checking tools: comparison and analysis.

Websites	Content	Topics	News Labels
FactCheck.org	Debates, news, interviews, TV ads	Multidomain: politics, healthcare news, immigration, climate change, etc.	False, no evidence, true
PolitiFact.com	Statements	American politics	Pants on fire, false, mostly false, half true, mostly true, true
Snopes.com	Videos and news articles	Multidomain: literature, economics, technology, crimes, scams, etc.	True, mostly true, mixture, mostly false, false, unproven, outdated, miscaptioned, correct attribution, misattributed, scam, legend
AFP Fact-check	Online rumours and claims	Environment, science, politics, health	Altered, fake, misleading
Fact checker	Claims and statements	American politics	One pinocchio, two pinocchio, three pinocchio, four pinocchio, the geppetto checkmark, an upside-down pinocchio, verdict pending
FactCheckHub	Claims, fake news and rumours	Science, politics, healthcare, environment	False, misleading
OpenSecrets.org	Biased political policies, fake news, statements, news articles, fraudulent web comments	American politics	Misleading, fake

Table 6. *Cont.*

Websites	Content	Topics	News Labels
Hoaxy.iuni.iu.edu	Tweets, claims	Multidomain: healthcare, economics, technology, crimes, scams, etc.	Altered, fake, misleading
Factmata.com	Claims	Multidomain: crimes, technology, scams, politics, economics, etc.	False, no evidence, true
TruthOrFiction.com	e-rumours and rumours, scam emails	Myths, politics	Fake
Telugpost.com	Fake news, claims, spam emails, fake posts	Multidomain: politics, healthcare, economics, religion, entertainment, technology, weather, etc.	Misleading, fake
The Quint	Fake news and posts	Multidomain: healthcare, politics, economics, entertainment, religion, weather, technology, etc.	True, misleading, fake
Vishwasnews.net	Fake news, claims, clickbait	Multidomain: politics, healthcare, religion, economics, entertainment, etc.	True, misleading, fake
Soch Fact-check	Claims, online posts, fake news	Politics, technology, entertainment, healthcare, etc.	Fake, misleading

- **Fake News Detector:** It is an open-source project used to identify FN. The users can mark news as clickbait, strongly biased, or FN. The other users of the fake news detector can see and add flags to already flagged news. The flagged items are saved in the repository and made available to Robhino, an ML robot, which classifies the news automatically as clickbait, fake news, or highly biased news based on human input.
- **SurfSafe:** FN can be analysed using different methods, including textual and visual analyses. The developers of SurfSafe evaluate a multimodal dataset collected from 100 fact-checking and renowned news websites for FN classification. The developed plug-in verifies the images in the curated dataset. The new image is compared to already present images in the dataset. The modified, fake images, or images used in a misleading context are classified as FN.
- **Trusted News add-on:** The tool assists the users in identifying FN and misleading claims. It is developed in cooperation with MetaCertProtocol. The users may assess the legitimacy of online content, which is classified into three categories. A broader range of outputs classify the web content as satirical, biased, malicious, trustworthy, untrustworthy, clickbait, and unknown.
- **FiB:** FiB considers post-production and post-distribution factors because content dissemination and creation are equally significant. The legitimacy of the post is validated using AI. The trust score is assigned to each verified post after source verification, image recognition, and keyword extraction. Additionally, FiB adds and validates facts against each misleading and fake post.
- **BS-Detector2:** It is available for both Mozilla and Chrome as a plugin. Each link on a webpage that leads to an unreliable source is compared to links to a list of domains. The domains are categorised as FN, satire, proceed with caution, clickbait, strongly biased, conspiracy theory, etc.
- **Fake News Guard:** The tool verifies the web links posted on Facebook and other accessed web pages. It is available as a browser extension. The tool provides fact-checked content using AI techniques and network analysis. However, there is a lack of details regarding how this tool operates.
- **BotOrNot:** The open-source service is provided to users to classify Twitter accounts. The accounts that show social bot-related features are assigned a score. Various

classification features are employed to classify Twitter accounts' content, metadata, and interaction patterns. Six classification features are related to user, network, friend, content, temporal, network, and sentiment.

- **TrustyTweet:** The browser plug-in recommended for Twitter users evaluates and improves media literacy. The emphasis is shifted from labelling FN to assisting people in making their judgments by offering clear and unbiased guidance when dealing with FN. Previously identified and effective potential FN indicators are used to guide the users.
- **Decodex:** The FND tool issues FN alerts to enlighten the user through the content classification labels satire, info, and no information.
- **LiT.RL News Verification:** The tool investigates and evaluates the news features on websites. NLP and SVM-based textual data analysis and automatic classification are the primary features of this tool. FN on the website is recognised and classified as satirical, clickbait, and fabricated news.
- **TweetCred:** The plug-in evaluates tweets' veracity using a supervised ranking classification trained on more than 45 features. Each tweet on the timeline is rated based on its legitimacy. The tool assigns a credibility score for 5.4 million tweets installed 1127 times over three months.

5.3. Crowdsourced Fact-Checking

Effective fact-checking involves various steps and aspects to evaluate the news veracity [219]. Recently, the crowdsourced fact-checking approach has been used and considered promising for identifying misleading information accurately and quickly [220]. This allows democratisation due to wide community engagement with reliability, scalability and robustness. This approach can potentially present a solution to curb the threat of online misinformation and disinformation. However, certain aspects hinder its widespread applicability and acceptance. Quality assurance, accountability, user expertise and bias [85] remain the main complexities and undermine the credibility of the crowdsourced fact-checking approach.

5.4. FND Based on Social Practices

There are certain limitations to fact-checking websites and tools. The most significant are a limited number of fact-checked FN, domains, regions, and languages. Extensive search and expertise are required to verify the misleading and forged multimedia content, and the process is time-consuming with delayed outcomes. Additionally, there is a significant disparity and difference in opinion about fact-checking services, and most fact-checking websites disagree and disapprove of their counterparts. It may further confuse users. Therefore, online users must improve their ability to distinguish between authentic and fabricated news. Researchers and practitioners are driven to develop systematic models for adaptive, automated, and comprehensive prediction approaches for FND. The review gleans the practical, social theories-based approaches for FND. These approaches are "news content-based", "social context-based", and "creator-based", and present a promising solution.

- **Social context-based analysis:** An alternative realistic approach is to record the cultural context of news sources. The examples entail evaluating the news period and connected sources and establishing the credibility of news sources. The study implements the link structure pattern of news-related websites as a data source for FND [221].
- **News content-based analysis:** The users face a sheer volume of shared FN online with attention-grabbing headlines (clickbait). The users cannot distinguish between fake and accurate news [222]. Social theories assist users in identifying FN articles. The users should not fall prey to the news headlines. FN and clickbait entice users to visit a webpage. However, they may contradict the assertions of the news. Therefore, reading the entire article is deemed necessary [223]. FN writers include various

facts, including analytical, research material, and expert knowledge frequently with supporting citations, assertions and links. Reviewing the articles in depth indicates to the users legitimacy and truthfulness [222]. FN addresses the users' concerns and apprehension on purpose. Additionally, FN items posted on OSNs may contain delicate subject matter, and users must continuously question whether the material appears appealing or funny so as to be true. Such information may indicate impending calamity, such as an earthquake, flood, etc., or information about an illness.

- **Creator-based analysis:** The researchers determine that FN can be detected using characteristics based on news content. Therefore, evaluating news sources rather than FN and fake claims provides better opportunities to users for FND. The users can identify potentially fake content and clickbait headlines based on the social information related to content. Additionally, the literature confirms that users cannot distinguish between fake and true content [224]. Users can recognise a fraudulent website by its web address, which may have suspicious tokens and unusual domain names. The "About Us" or "Disclaimer" information of a website also provides helpful information and proves the legitimacy of the website [225]. Additional techniques for FND are news content evaluation, creator and consumer evaluation, and social context evaluation.

5.5. FND Based on Related Components

5.6. User and Creator Analysis

Analysing fake OSN profiles involves numerous efforts and steps. Due to the prevalence of unverified communications, internet users can use tools to assess the legitimacy of online connections [226]. On OSNs, fake profiles spread misinformation on purpose to influence users' behaviour [227]. Therefore, users and FN spreaders are analysed to identify FN. On OSNs, fake accounts behave differently from verified users due to their unique characteristics. The further investigation for FND can be divided as follows.

5.6.1. OSN Profile Analysis

The native language of users provides identity information. It also includes the user's registration date, geographical information about the account, profile authorisation, the user's number of tweets or posts, etc. [227]. Fake OSN accounts can also be identified using consumer profile research and the status of an online service [228].

5.6.2. Post-Sharing Analysis

Longitudinal activity and signals that fit into a Poisson distribution determine the social community profile of OSN users. Time is another significant factor, such as the average sharing time between two articles, response time, sharing and discussion frequency, and time taken by the programs and users in managing the process. During specific periods of the day, abnormal profiles, such as social bots and cyborgs, are substantially more active [229] compared to genuine users [230]. Fake accounts spread misleading news and information disproportionately and concealed their regional identities while participating in discussions and responses to events during the 2016 United States presidential election [230].

5.6.3. Credible OSN Account Information

These contacts measure a formula for determining profile popularity based on the number of followers and connections. Another useful feature to discern between fake and legitimate profiles is the number of followers and friends. For a genuine profile, the number of fans on an OSN is typically close to the number of friends. Illegitimate bots have fewer fans and more contacts [70].

5.6.4. User Sentiment Analysis

Additionally, unverified accounts are detected using sentiment criteria [228]. Fake profiles generate fake information and the emotional reactions in the FN perplex authorised users [231]. The sentiments, mindsets and perspectives expressed on OSNs can be comprehended through emotion analysis. FN and forged information creators methodically generate a psychological word stream intentionally [232]. The expressive rating analysis is sentiment based. The study [233] determined the happiness index. The research determined intensity, polarity, and emotion rating [234], as well as various emotion-related characteristics [234].

5.7. Unimodal (Single-Modal) FND

FND is based on three major features, i.e., news content-related, user-related, and social context-related features. Single-modal FND evaluates a single feature. Style- and visual-based features are related to news content. The evaluation of words, sentences, and news items yields the characteristics of news content. These features are also named linguistic- and syntactic-based features, respectively [235]. Information on user profiles, credibility, behaviour, and usage patterns are examples of user features. Lastly, the news circulation method, user's reaction, and interaction on OSNs are categorised as social context features. Social context features also include network-based, distributed, and temporal features. Single-Modal FND methods focus on news content-related features. The accuracy of the adopted method for FND depends on considering multidimensional features instead of following a single-modal approach [236]. Malicious content creators use different writing styles to entice users to read online news on OSNs, blogs, and news websites. Therefore, the auxiliary information, i.e., relevant features-related information, is significant.

5.8. Multimodal FND

Multimodal FND concentrates on feature combinations [237]. The existing literature on multimodal FND provides insights into two approaches. The first is content-based features that include evaluating the content, images, and videos. The second is a collective analysis of user profiles, i.e., user-related information with news-related features (news content, style, etc.) [238]. The multimodal FND process is illustrated in Figure 11.

Textual and Visual Features Detection

The previous work on the multimodal FND was executed with relevant datasets. The approaches for multimodal FND to classify FN and extract features are presented in this review. The author implements a CNN algorithm for textual and multimedia data. Bidirectional Encoder Representations from Transformers (BERT) is used for unimodal data analysis. Different types of FN, such as forged content, misleading claims, satire, etc., are classified using the approaches. The method achieves high accuracy for the multimodal dataset, Fakeddit. The author proposes a multimodal model for FND [239]. The feature selection is based on two categories of visual attributes, i.e., visual features and statistical characteristics of news. Five visual features reveal concealed distributions in disseminated news content, and seven statistical features detect image distributions. The multimedia dataset Sina Weibo is evaluated using different ML-based classifiers, i.e., RF, LR, K-start and SVM. The authors demonstrate the effectiveness of combined features for multimodal FND on a real-world dataset [240]. Semantic, textual and visual features are consolidated for the final predictions. A semantic and contextual analysis is pursued using BERT, and the VGG-16 model is used for visual features. The cosine similarity between the news heading and multimedia tag embeddings is used to determine the text-image similarity referred to as the semantic representation. Two studies present multimodal FND using DL models [241–243]. BERT is used for content-related features, and VGG-19 is implemented for multimedia features for the proposed SpotFake system [244]. SpotFake+, the updated version of the multimodal system, is based on Transfer Learning (TL) for feature extraction. The pre-trained model XLNet is used for content-related features, and VGG-19 is implemented for multimedia features. The final predictions are based on the combined features. The authors divide the proposed model into three main sub-components [242]. The first is a textual feature extractor that employs sentiment analysis to investigate important information from news content. The second component collects visual features from the pre-processed posts using a segmentation algorithm. The best features from the news content and visual feature representations are extracted using a cultural algorithm. The FN detector classifies true and FN. The cultural algorithm for the best feature extraction shows the best performance. Three CNN architectures, i.e., ResNet50, VGG-19 and InceptionV3, are implemented for visual feature extraction [245]. Bidirectional LSTM is employed for textual feature extraction. FN is classified based on the combined analysis of content and multimedia features. The shared task, Facity2, is held for multimodal news dataset evaluation [237]. The information requires a comparison-based approach for the assigned task, which pairs OSN claims with documents to classify textual and multimedia data based on multimodal features. The DeBERTa-based model is used for content classification, and the SOTA CLIP and Swinv2 techniques are implemented for images. Three modalities, i.e., content and multimedia-related features with textual and visual information, are proposed for FND [246]. ML-based approaches are used for news classification. Predictive analysis is performed on both characteristics, i.e., textual and visual, to determine their relativity to FN.

5.9. FND Algorithms

Different algorithms are used for content- and context-based FND, sentimental analysis, etc. The significant approaches include ML classifiers, and stacking ML classifier-based techniques, DL models, pre-trained transformer-based models, and mixed models, which are implemented to increase the classification and prediction efficacy of multiple submodels. We have included basic information related to each group of algorithms and an additional short description. Various approaches require labelled data, while others process unlabelled data. We present the FND process, where the learning algorithms predict the final class of a news item (input). Content-based FND is related to the class prediction of FN. Different classification metrics, F1-score, accuracy, precision, recall, MCC-values, etc., are used to demonstrate the final class predictions. The stacking or ensembling technique increases the prediction performance of the selected algorithms. The stacking model consolidates the predictions made by each model and presents the final prediction result. Stacking, bagging, and boosting are the chief ensembling methods. ML-, DL-, and NLP-based approaches can be used for content-based FND as depicted in Figure 12.

5.10. FND: ML-Based Approaches

A subfield of AI and computer science is ML, which focuses on data and algorithms to simulate the human learning process and progressively increase the overall accuracy. These methods improve performance on a set of tasks by using data. Different component classifications extract significant digital communications information, subsequently applied to develop learning methodologies. Three major categories of ML data mining approaches include unsupervised, semi-supervised, and supervised models as shown in Figure 13. Traditional ML parses data, and after training on the dataset, it makes predictions. ML algorithms are trained robustly and easy to implement with low-resource requirements and show better results on limited-sized datasets. DL-based approaches have shown great success in speech recognition and visual object recognition-related tasks [12,34]. Works in the literature have used an ensemble-based approach to extract various language-based characteristics from the news articles using Linguistic Inquiry and Word Count (LIWC) [89]. The proposed ML-based approach for FND uses an optimisation process with the Genetic Algorithm (GA) [247].

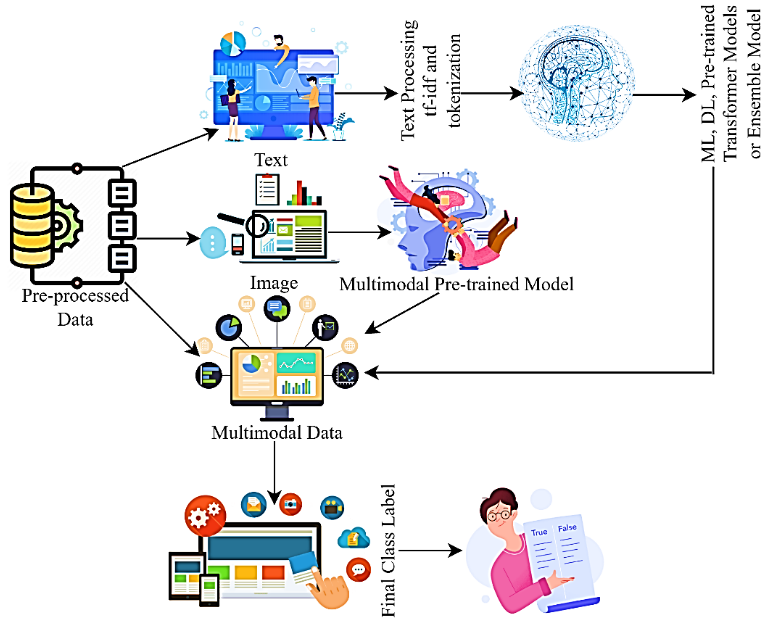

Figure 12. Multimodal Fake News Detection process.

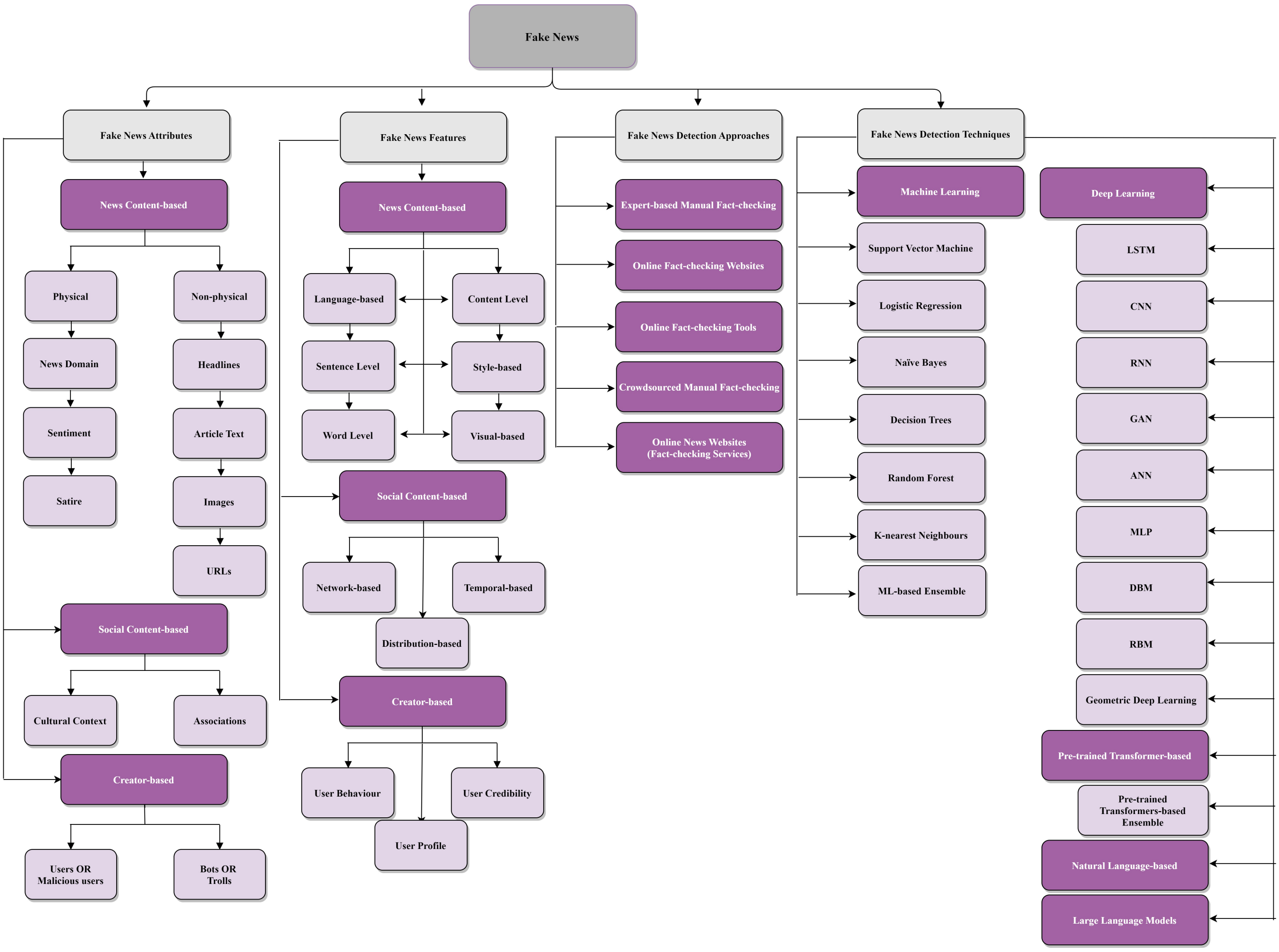

Figure 13. Algorithm classification for existing Fake News Detection.

Supervised Learning

Supervised learning is a higher-level class of ML algorithms which requires labelled data for training and predicting accurate results. The existing research has determined the effectiveness of supervised ML techniques for classifying misleading, forged news, clickbait, hoaxes, etc. [248]. Classification and regression signify two major tasks of supervised ML. Classification is used widely for class predictions of discrete values (class labels), such as true, false, misleading, etc., for FND, male and female for gender classification, malicious or non-malicious for UAV traffic, etc. Contrarily, regression-based methods predict the continuous values (continuous quantity), such as estimated salary, price, age, etc. Different supervised ML techniques are Support Vector Machine (SVM), Random Forest, Logistic Regression, K-nearest Neighbour, Multilayer Perceptron, Decision Tree, and Passive Aggressive Classifier [226].

- **Support Vector Machine (SVM):** SVM is used for identification, classification and regression. The labelled data are fed to the models to acquire the results. SVM generates the best results for identifying FN [249]. The study shows the best performance with accuracy using SVM [190].
- **Logistic Regression (LR):** It is necessary to anticipate a classification value whenever the LR approach is used. The method is implemented to make a reasonable assumption or to present the results as correct or incorrect. The study shows that the approach is used to evaluate data accuracy [250]. There are three categories of LR, i.e., Multinomial LR, Binary LR and Ordinal LR. Term Frequency-Inverse Document Frequency (TF-IDF) is used with LR in the study for better results [251].
- **Naïve Bayes:** The accuracy or deceptiveness of the content is determined using Naïve Bayes [252]. There are different types of Naïve Bayes classifiers: Bernoulli, Gaussian, and Multinomial Naïve Bayes classifiers.
- **Decision Trees (DTs):** The dataset is divided into multiple sub-categories, and the supervised ML-based method DTs are used for FND [253]. Bootstrap aggregated trees (bagged trees) and boosted trees are two types of DTs. FND is possible using textual features in addition to tweets and user-level features.
- **Random Forest (RF):** RF classifier is used for FND. The approach implements many DTs on different dataset subgroups. The final result is an aggregated average value of the prediction results. This method improves the overall prediction accuracy. Different ML-based approaches are employed for FND [254]. The ISOT dataset is evaluated using the RF-based method and shows the best accuracy [255].
- **K-Nearest Neighbours:** The approach is followed to address classification-related tasks. The information is collected for each entity, and based on the existing characteristics and similarities, new entities are classified. The study uses the approach for FND [256].
- **Gradient Boosting:** The prediction model contains a stacking model of weak classification models, such as DTs. The approach allows the boosting of differentiable loss functions and constructs an additive model in a forward stage-wise manner [257]. N regression trees are fitted on the loss function's negative gradient at each stage.
- **Extreme Gradient Boosting:** It represents the quickest way to use gradient-boosted trees. It generates a new branch by contemplating the potential loss for all feasible splits [222]. The distribution of features among all the data points in a leaf and algorithm is considered, and the inefficacy is addressed using this knowledge to condense the space of potential feature splits.
- **Bagging classifier:** An ensemble meta-estimator is applied to base classifiers, such as DTs, to select random subsets of the original dataset [258]. The final class prediction is selected based on the prediction made by each model for the ensemble. The voting techniques or averaging method select the final class prediction.
- **AdaBoost:** The adaptive ensemble model adjusts weak learners to improve the prediction performance for classifying the inaccurate categorised samples [259]. Each selected model for the ensemble may have lower prediction performance, but the

final model becomes an effective learner model, given that each learner exhibits better performance than random guessing.

5.11. Unsupervised Learning

Unsupervised learning-based methods evaluate unlabelled samples (raw data). The technique clusters data and detects hidden data patterns without human intervention. However, the labelled dataset plays a significant role in training a model for achieving high accuracy for FND. As mentioned earlier in Section 4, curating a labelled large-sized and multidomain dataset is challenging. FN on OSNs and online websites is massive in quantity, distorted, unstructured, and imprecise [66]. A substantial amount of deceptive and misleading content is created and shared on OSNs, blogs, and online websites serving different goals, targeting specific readers and audiences, and exhibiting linguistic abnormalities [260]. Therefore, curating labelled datasets based on facts and figures is difficult. FND is a real-world problem which can be addressed using an unsupervised learning approach. However, relatively less research has focused on FN disseminated uncontrollably. The existing research has mainly focused on latent semantic and sentiment analysis. Ref. [261] assesses fake reviews using unsupervised similarity. The suggested methodology provides higher accuracy for identifying virtually identical online ratings, implementing word similarity, and word-order similarity in conjunction. The literature proposes an unsupervised sentiment analysis method for OSN images [262].

- **Semantic Similarity Analysis:** The author suggests the semantic similarity analysis of FN similar to authentic news [261]. Due to a lack of pertinent data and inventiveness, FN frequently reuses the content of previously published news stories, and, often, outdated news is shared as well [263]. FN creators can change a few words in a comment to deceive a fictitious online reviewer. Therefore, semantic similarity evaluation is an effective technique to detect FN, forged content, and misleading reviews.
- **Outlier Analysis:** Outlier analysis is the process of gathering data to detect malicious behaviour. The outlier analysis can expose fabricated information and illegitimate creators, length measurements and density-based techniques through the appropriate statistical metrics [264].
- **Cluster Analysis:** Cluster analysis evaluates unlabelled online data. Clustering algorithms, which can also create classifiers for data collection, can be classified by maximising intraclass similarities and minimising interclass similarities [265]. FND using cluster analysis makes it possible to identify particular homogeneous classifications of news and writers.

5.11.1. Unsupervised News Embedding

The process of acquiring distributed representations of unprocessed data is known as embedding, and it is a crucial stage in NLP. The numerical embeddings are a source of additional analysis in FND. Different embedding systems can detect data properties from different relevant perspectives. The effectiveness of the FND process and misleading information identification depends on selecting an embedding method to determine the true nature of the news. The prominent unsupervised-based embedding method is word2vec [266]. The study shows that FastText creates supervised and unsupervised learning algorithms to produce word vector representation [267]. Using an unsupervised model, Sent2Vec, a general-purpose sentence embedding, may be learned [268]. Each document is represented as a vector by the Doc2Vec model [269].

5.11.2. Different Ensemble Learning Approaches for FND

The author proposes a novel DL-based approach for FND [270]. Various training models are fed pro-processed news items. Data tokenisation, text and grammar analysis, and LIWC are performed to acquire bigrams and uni-grams [271]. LSTM, depth LSTM, n-gram CNN, and LIWC CNN models are used for the proposed ensemble learning model. The weights of the proposed ensemble learning model are converged and optimised using the Self-Adaptive

Harmony Search (SAHS) algorithm to achieve the best prediction performance. Multidomain oriented datasets are used to evaluate the cross-domain intractability problem. The study presents various methods for FND based on the extracted attributes from the news content without using any additional content and context-related information [272]. Stylometric features and text-based vector representations are used for the suggested ensemble model. Additionally, bagging, boosting, and voting features are selected for the ensemble learning model. In addition to the news content, media content is used in designing the model. The study proposes an ML-based classifier ensemble model for FND. The model is implemented to extract significant FN dataset features. DT, RF, and Extra Tree classifiers that show the best overall accuracy are stacked for the assigned task [273]. The authors have determined the efficacy of ML algorithms for FND. An FND system based on ensemble voting classifiers is used for FN classification. The three best performing ML algorithms out of the eleven most used ML models are selected for the proposed FND system [274].

5.12. FND: DL-Based Approaches

DL—a sub-category of ML—processes complicated problems like a human brain and presents the probable solutions [275]. The author implements DL techniques to identify text-related issues, such as FND and spam detection. DL provides an impetus for FND [48]. The literature suggests that neural networks are a significant method for FND [73]. Dechter was the first to apply the DL in the ML field and employ artificial neural networks for Boolean threshold prediction. The implementation of DL in AI research is significant and used for a wide range of applications, including personality mining, asset allocation, anomaly detection, speech recognition, computer vision and NLP. It is used more frequently to aid decision-making through data analysis and trend identification. The innovative approach also aids in expanding the study field, improving learning execution and streamlining the measurement procedure [260]. Over the past few decades, many studies have been proposed to solve different content-related challenges on OSNs, such as FN, misinformation and disinformation, anomaly detection, etc. The research community explores innovative techniques and fields of study to address the identified research gaps. DL has gained the popularity and approval of researchers [276]. Other neural networks, such as Long Short-Term Memory (LSTM), Recurrent Neural Networks (RNNs), and CNNs, are developed to gain valuable information and expertise from various implementations. DL approaches identify original data representations automatically compared to typical ML methods, which require manual feature detection [277]. DL-based techniques have shown the best performance for voice and visual object recognition [12,277]. The existing works in the literature use Gated Recurrent Units (GRUs) for FND [268], and LSTM-based FND is presented by [278]. Bidirectional Recurrent Neural Networks (BRNNs) enhance prediction accuracy [279].

5.12.1. Deep Neural Networks (DNNs)

DNN is the widely used DL network. It is an input- and output-layered neural network with a single hidden layer. This feed-forward network has no loops and transfers data from the input to the output layer. The information processing and communication nodes in biological systems serve as the basis of DNN. The input data are directed to various network layers, and the neural network nodes are comparable to neurons in the brain. The system presents the processed information in a specific manner. The network is instructed using the back-propagation technique. Additionally, the model is trained to understand data characteristics quite effectively.

5.12.2. Long Short-Term Memory (LSTM)

LSTM is a sequential neural network that permits information persistence [280]. This type of RNN can solve the vanishing gradient issue encountered by a typical RNN. It deals with sequential data, such as time series, audio, and text. It is used for various tasks, such as language translation, time series prediction, and speech recognition because the networks can learn long-term dependencies in sequential data and remember the past information [281]. By employing LSTM units, which are made up of several gates and are in charge of keeping track of a concealed cell state, LSTM can avoid the vanishing gradient problem. Because of this, LSTM units can recall data from a longer period than traditional recurrent units [276]. It is a crucial component of NLP since the past information affects the current information. In the architectural history, there are two LSTM layers: one for the forward data feed and one for the backward data feed.

5.12.3. Convolutional Neural Network (CNN)

The convolution technique is used by CNN to produce outputs through matrix multiplication in subsequent training. CNN focuses on data processing to gain knowledge and data patterns. NLP uses word vectors to represent news items and sentences [165]. The word vectors are used to train a CNN. Training is possible because of kernel and filter size specifications. A CNN has a multidimensional capacity to perform the assigned tasks. The layered network of neurons processes inputs and outputs. CNN is a feed-forward network paradigm that performs object identification and picture analysis. CNNs are very helpful for recognising classes, objects, and categories in photos by looking for patterns in the images. They can be useful for signal classification, time series, and audio data. A neural network has three layers—a convolutional layer, a detector layer, and a pooling layer. As previously mentioned, the convolution layer produces a map with convoluted features. Feature maps exhibit observable non-linear components, such as identified features in the detector layer. The pooling layer offers both an output and a reduction in the preceding information. The two main applications of this approach are data sizing and data training. CNN is used for text and image feature detection in [282]. The text-CNN-based approach is implemented for FND in [283] and health-related problems in [284].

5.12.4. Recurrent Neural Network (RNN)

RNN—a feed-forward network that uses sequential data processing for learning as sequential processing—can keep track of previous events. Recurrent processes employ the outcome from a single time step as the input for the next time step. The findings of the previous step are stored in a temporary variable. RNN identifies interdependent data patterns from training data. The model is trained using backpropagation. The ANN model is a subset of RNN and uses recurrent loops for sequential data analysis. It is used for various applications, such as sentiment analysis and speech recognition. RNN-based models have higher memory and produce results based on the previous inputs. The model comprehends human language and responds accordingly. RNNs are employed in Siri on the Apple iPhone and Amazon Echo. It recollects prior inputs and applies the same settings to generate the output. RNN takes advantage of the embedded structure and focuses on embedded architecture in the data sequence for acquiring useful knowledge. The model has the advantage of collecting more contextual data. RNNs can identify sequential patterns in specific data, i.e., for FND [285]. In particular, subject classification [286], question-answering [287], sentiment analysis [287], and language translation [288] have shown their significance in NLP.

Generative Adversarial Network (GAN): The GAN technique creates new statistical data comparable to the training dataset once it is trained. The literature developed a GAN model to improve the automated rumour identification performance and to detect indistinct or inconsistent voice input characteristics [289]. Previous research has shown that the most common cause of widespread rumours is the intentional dissemination of misinformation and disinformation to promote consensus through rumour-related news.

Artificial Neural Network (ANN): ANN evaluates and identifies the relationship between input and output variables. An ANN structure uses the input dataset to predict output variables [290]. ANNs are used to classify satellite image data. It is quite similar to a human neuron. This feed-forward model has hidden layers, input layers, and output layers. The input layer serves as an interface to receive data and communicate with the hidden layer. In conjunction with the input layer, the outcome from the hidden layer is sent to the output layer. The neural network is fed different inputs or outputs to compare the real output and ANN output.

Multilayer Perceptron (MLP): MLP is a type of ANN that replicates the human brain function. Classification, regression, and clustering are addressed using an ANN that learns from training data and recalls and implements the learned knowledge. Due to its functionality, MLP is also called a black box. Numerous calculations are involved using an MLP [291]. Synaptic weights connected to neurons in the hidden layers are presented as the inputs through the input layer. The functionality of the hidden layers of an MLP depends on data complexity. Each subsequent hidden layer is related to the previous layer via synaptic weights. Each layer depends on the output of its preceding layer. Once the input is fully processed, the prediction output is generated at the output layer.

Deep Boltzmann Machine (DBM): A DBM is known as a binary pair-wise Markov random region because it contains numerous hidden random variables. A network of symmetrically connected Stochastic Binary Units (SBUs) can be used to identify malicious activity. The author identifies FN using a multimodal benchmark dataset and employs a multimodal DL model based on DBM to segregate spoken queries [292]. In a DBM, units within each layer are not linked, but layers are interlinked.

Restricted Boltzmann Machine (RBM): A particular class of neural network is a generative stochastic ANN (RBM). It may identify a possibility that spreads across the set of inputs. The limited Boltzmann machine, a variation of the general Boltzmann machine, supports learning. Connections between hidden units within a layer are, therefore, not allowed [293]. RBMs can be stacked to efficiently train several hidden units. RBMs can be used for autonomously extracted spam detection properties.

5.12.5. Geometric Deep Learning

Michael Bronstein created a DL technique called geometric deep learning that involves embedding a geometric data understanding as an inductive bias DL model. Graphical Deep Learning (GDL) has achieved substantial success both theoretically and in real-world applications [294]. The Boundary Samples Mining Module (GSMM) and the Future Scaling Module (FSM) are defined [295]. Gabriel's graph recommends implementing the BSMM to derive boundary values and the probable link relationship. The methods address the imbalance problems. The study shows that using the geometric DL technique, it is possible to predict creating a graph based on an original and intricate retweets network using the fewest computing resources with high accuracy [296].

5.13. NLP Techniques for FND

Domain-specific features were the main focus of the NNs before the developed NLP approaches. Traditional methods for FND focus on linguistic characteristics, such as syntactic and lexical patterns [231,297]. However, with the advancements in DL, researchers have selected more complex techniques, i.e., RRNs [298], CNN [299], and pre-trained transformer-based models [300], to derive contextual and semantic features from the news content. These models also improve the detection accuracy. NLP is an emerging field of ML that learns, analyses, manipulates and integrates human languages. NLP has provided an impetus for new research and presents future research directions with promising results and higher accuracy [301]. Speech recognition, sentiment analysis, pragmatic analysis, machine translation, disclosure integration, automatic message digest, chatbots, automatic question and answer generation, intelligence, and text classification are some of the widely adopted applications of NLP. The NLP framework is presented in Figure 14. The tasks, such as dependency parsing, parts-of-speech tagging, and conference resolution, are included in the pipeline for some sophisticated NLP techniques [302]. Automatic sentiment detection and opinion mining from the text have developed tremendously because of NLP algorithms [303].

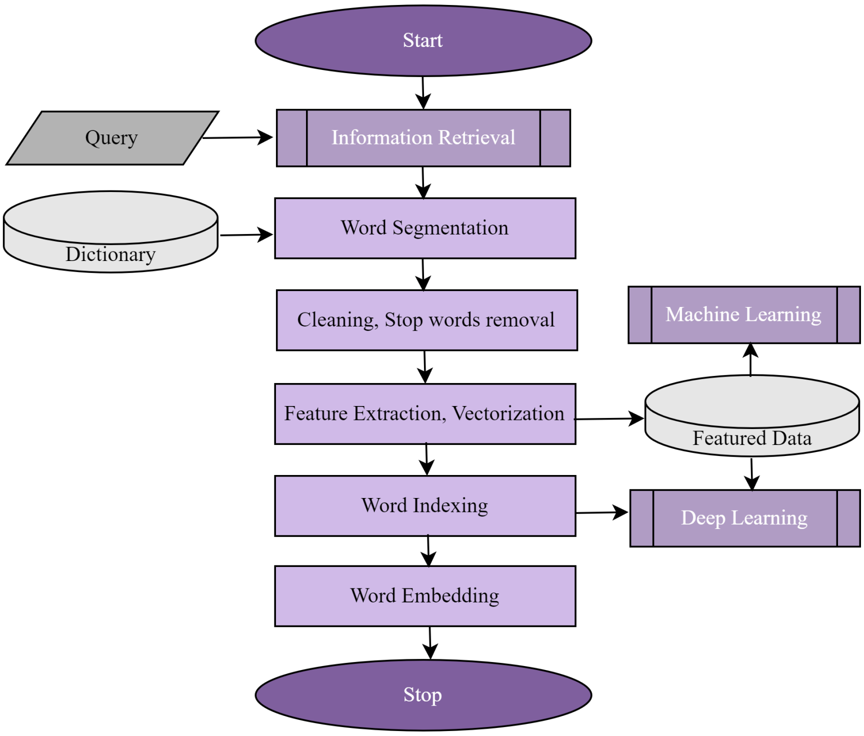

Figure 14. NLP FND framework.

Due to extensive studies on FND, benchmark datasets and methods have been developed [302]. Unified NN architecture and methods can be used for multidimensional NLP-related tasks [304]. They use pre-processing, word embedding, and feature extraction approaches. Data pre-processing represents cryptic attributes, data cleaning, stop-words removal, tokenisation, and the creation of intricate structures with attributes, which is the first stage in the development of FND models. Different visualisation approaches are helpful for data pre-processing. It reduces the required computing resources and computational time while addressing noisy data. In the second step, word vectorising, text or words are mapped to a vector list. The process also enhances the overall prediction performance of the models for the assigned task. ML-based models frequently use TF-IDF and bag of words (BoW) [305,306] for FND. Due to their capacity to train on larger datasets, word-embedding models, such as word2vec and GloVe, have been used for FND models. Large memory and computing resources can investigate a large number of variables. Classification methods boost overfitted and inadequate samples. The existing literature implements stylometric features in OSN content analysis. The feature extraction process produces combinations of variables to overcome these issues for data classification with accurate precision. FND models use social context features to extract opposite attributes from the news content [204]. The n-gram algorithm processes news content and produces words and characters from multidimensional n-gram vectors [34,307–309]. The n-gram vectors are grouped to create a feature vector for each piece of content. Linguistic feature extraction has various feature classes, including user credibility, style-related features, quantity features, readability index, and psycho-linguistic features. These features determine the prediction performance of FN [190]. Word vectors are produced via word embedding for the subsequent tasks [310–313]. However, the construction of word vectors from a sizable dataset is challenging. This review highlights various FN-related tasks with their characteristics and challenges to be addressed in future research.

Pre-Trained Transformer Models for FND

A pre-trained model is trained on very large-sized datasets to perform the assigned tasks [314]. DL models are expressions of neural algorithms that resemble the brain and detect patterns or produce appropriate data-based classification predictions. The wide-scale application of pre-trained models in industries is based on modifications of the models according to the requirements. Instead of starting from scratch when creating an AI model, developers can leverage pre-trained models and modify them to suit their needs [315]. The learning process involves passing through various data layers and analysing goal-oriented data characteristics. Starting with the basic representative data layers, complex data patterns required for evaluation are developed. Calculated probabilities provide these attributes with varied degrees of relevance [316]. Developers need large-sized datasets, frequently containing billions of data samples for pre-trained models. The compromised dataset affects the model performance extensively. Data acquisition is expensive and challenging. Higher prediction performance and model deployment result from using a high-quality pre-trained model with a significant number of correct representative weights. The modification of weights and extensive data addition adjust or modify the model. AI applications are developed promptly using pre-trained models. Moreover, the developers' concerns about loads of input data and probabilities analysis for dense layers are addressed [317]. Time, money, and effort are saved using weights, which are pre-computed probabilistic representations [318]. An encoder and a decoder formulate two main sub-components of the transformer architecture. Self-attention (the essence of pre-trained models) is used to compute an input representation without recurrence or convolution in the encoder. An encoder is a stack of identical layers that comprises a multihead self-attention mechanism and a fully linked feed-forward sub-layer. The architecture of the decoder is similar, with the encoder output covered by an additional intermediate layer for disguised attention [319]. NLP, computer vision, healthcare, speech AI, cybersecurity, and art and creative workflows are among the top fields where pre-trained models are developing AI [320]. BERT and the Generative Pre-trained Transformer (GPT-n) series are two examples of effective large-scale pre-trained language models [321].

- **Bidirectional Encoder Representations from Transformers (BERT):** The BERT model is different from other pre-trained language-based models. The model is trained to provide bidirectional language features. MLM conceals a portion of the input tokens and predicts the input tokens from left to right. Next Sentence Prediction (NSP) involves binary classification and determines two successive phrases [322]. The training corpora for BERT were the English Wikipedia (16 GB) and BooksCorpus. It outperformed the SOTA in eleven NLP-related tasks, including SWAG, SQuAD and GLUE, and demonstrated excellent results.
- **XLNet:** The generalised autoregressive model overcomes the weaknesses of BERT based on its autoregressive formulation by combining a bidirectional context and avoiding separate predictions [322]. The model introduces Permutated Language Modelling (PLM) and combines the autoregression advantages with MLM. The MASK token used during training disappears since MLM anticipates each masked token in a phrase. It causes a discrepancy between the fine-tuning and training process. PLM randomly permutes the input tokens. It predicts the target token and focuses on the preceding tokens in the permutation order. The traditional transformer structure has two additional elements in terms of architecture, i.e., target-aware and self-attention representations. It is pre-trained on BERT datasets, ClueWeb 2012-B4, Giga5, and Common Crawl5 data [322].
- **Robustly Optimised BERT Approach (RoBERTa):** The Robustly optimised BERT Pretraining Approach (RoBERTa) is a BERT-based pre-trained model. It addresses the limitations of BERT by improving the training process. The model is trained using different hyperparameters, i.e., different learning and batch rates [323]. The model is trained on a larger corpus, i.e., 160 GB of uncompressed text, the OPENWEBTEXT [324], CC-News3,

and STORIES [325]. RoBERTa outperformed several benchmarks, including SQuAD [326] and GLUE [304].

- **DistilBERT:** The size of the pre-trained model is reduced (40% compared to the original BERT) [327]; this is called DistilBERT, and the size reduction impacts its prediction performance. The student model (small) learns and replicates the teacher model (large) output distribution using a compression method known as knowledge distillation [328]. DistilBERT is a student model, and BERT is the teacher model in this scenario. It uses a triple loss for training, i.e., MLM, cosine embedding, and distillation loss. The training corpus of BERT is used to train DistilBERT. Being smaller and speedier compared to BERT, DistilBERT demonstrated prediction outcomes equivalent to those of BERT.
- **Multilingual BERT (mBERT):** mBERT [322] is a multilingual version of BERT [322]. It was originally trained in 104 different languages on a Wikipedia-sized dataset. The multilingual training provides prospects for researchers to implement the model for multilingual tasks in low-resourced languages. mBERT adheres to the development like mT5, and the implementation and architecture features of BERT.
- **A Lite BERT (ALBERT):** The model is designed to increase performance and speed with less memory (computing resources). The model employs factorised embedding parametrisation, and cross-layer parameter sharing is employed for reducing the parameters. ALBERT also introduces the Sentence Order Prediction (SOP) loss besides MLM loss [329]. It assists the model in predicting two successive sentences and detecting any changes in the sentence order. The model is trained on new data in addition to the BERT corpus.
- **Text-To-Text Transfer Transformer (T5):** The T5 pre-trained model [329] uses the basic transformer architecture proposed by [315], based on a basic encoder–decoder. The model is pre-trained on MLM with the aim of "span-corruption". It replaces successive input tokens with a masked token and trains the model to reconstruct these masked tokens. The scale of T5 is another distinctive feature, i.e., pre-trained using large-sized tokens ranging from 60 million to 11 billion parameters. Approximately 1 trillion data tokens were used to pre-train the model in addition to the C4 dataset (a collection of over 750 GB of unlabelled text) gathered from the open Common Crawl web scrape. It is widely used for NLP-related tasks [329]. T5 is trained to present the label output rather than a class index for classification tasks.
- **XLM:** BERT-based XLM uses better techniques for pre-training multilingual language models. The pre-trained model also attains the set objectives for cross-lingual pre-training [330]. It was originally trained on 100 languages from Wikipedia. There are many versions of XLM available, such as XLM-Large, etc.
- **XLM-RoBERTa:** It is an enhanced and multilingual version of XLM built on RoBERTa [330]. It was trained on Common Crawl in 100 different languages, and its main purpose is cross-lingual MLM modelling. An n-gram language model trained on Wikipedia was used to filter pages from Common Crawl to enhance the quality of the pre-training data [331].
- **BART:** The pre-trained model employs a typical machine translation or sequence-to-sequence architecture. It follows a left-to-right decoder (GPT) and a bidirectional encoder (BERT) approach. The original sentence sequence is randomly shuffled in the pre-training stage, and a new in-filling approach is used to replace long stretches of text with a single mask token [332]. Once the model is fine-tuned, the model exhibits better performance for comprehension-related tasks but is especially effective for text generation-related tasks. It obtains excellent outcomes for abstractive dialogue, summarisation and question-answering. The model showed the same performance as RoBERTa when trained on identical training resources to SQuAD and GLUE.
- **Multilingual BART (mBART):** The multilingual version of BART [332] was trained using sizable monolingual datasets in numerous languages [333]. The earlier techniques concentrated primarily on text reconstruction, encoder, and decoder, unlike the pre-training process of BART.

- **Multilingual Autoencoder that Retrieves and Generates (MARGE):** A multilingual encoder–decoder model was trained to restructure a document in one language using texts in other languages [332] The model was pre-trained on Wikipedia and CC-News data in 26 languages [323].
- **Efficiently Learning an Encoder that Classifies Token Replacements Accurately (ELECTRA):** The pre-trained model identifies replaced tokens (by a generator) in a sentence. The model requires fewer resources since it is a smaller model than BERT. ELECTRA follows the innovative pre-training method, i.e., trains the generator and discriminator transformer models [334]. The generator is trained as an MLM to substitute tokens. The discriminator determines the replaced tokens in the sequence. The transformer model shows outstanding results for multilingual tasks.
- **Other Pre-trained Models:** Transformers have become extremely popular in text generation and language-related tasks. The models discussed above are mostly employed for discriminative tasks. The relatively common transformer architectures Generative Pre-Training (GPT) [335], GPT-2 [336], GPT-3 [337], GPT-3.5, GPT-4 [338], and Large Language Model Meta AI (LLaMA) [339] are trained with a language model. These are also widely known as Large Language Models (LLMs). Although Natural Language Generation is their main area, their prominence in various AI fields is undeniable. The Longformer [340], Reformer [341] and Big Bird [342] were developed to handle extremely lengthy sequences (highly valuable for attention mechanism and increase input length exponentially) with high prediction performance and efficacy. A few other transformer models are Funnel [343], Pegasus [344], and CTRL [345]. The research community has released numerous pre-trained transformer models for different high-resourced languages, such as FlauBERT [322] for French, and Chinese BERT [346], a pre-trained model for Chinese. Additionally, the pre-trained models have provided an impetus for research and wide-scale applicability for multilingual and low-resourced language-based tasks, which were unattainable otherwise [335]. Different pre-trained transformers models are used for ensemble models for various tasks. The ensembling technique increases the overall prediction performance of the models for a target task. The general structure of an ensemble of pre-trained transformer-based models is given in Figure 15.

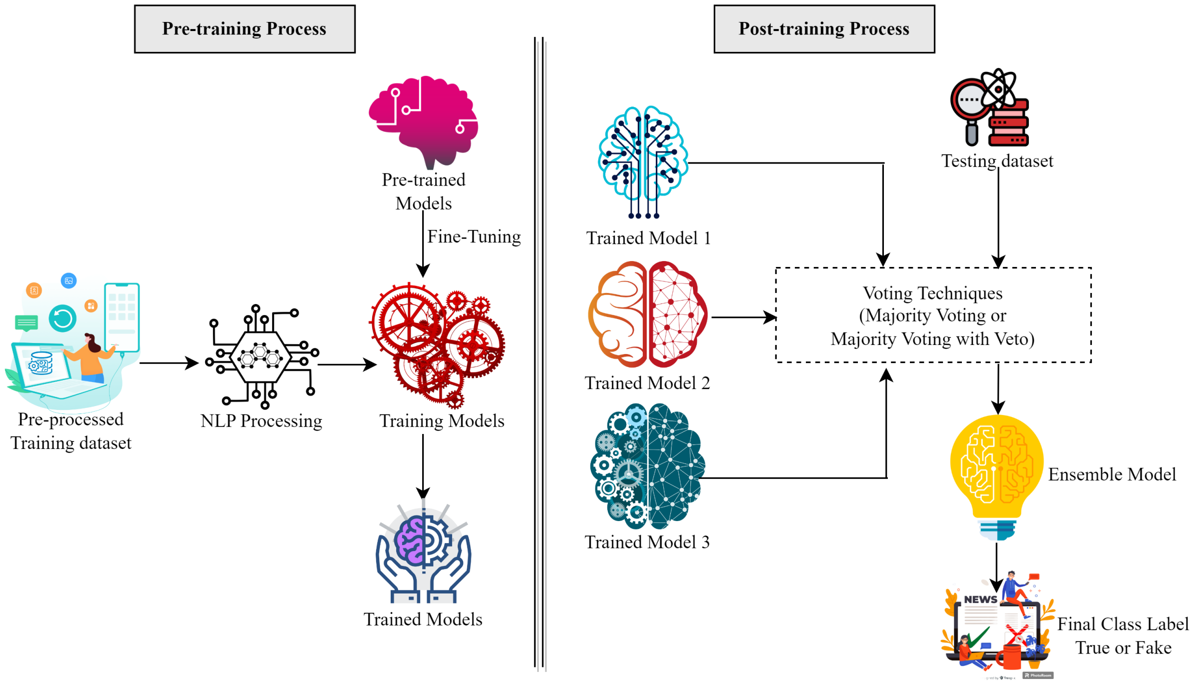

Figure 15. Pre-trained ensemble.

5.14. Large Language Models (LLMs) for FND

Large Language Models (LLMs) are AI-based algorithms that comprehend, summarise, generate, and predict new text using DL methods and extraordinarily large-sized datasets. LLMs are also closely related to the concept of generative AI. The models have been specially developed for automated solutions to content-related issues [347]. The users enter queries in natural language and produce results, and language models are widely adopted in NLP applications. AI technology is now widely relevant for FND since LLMs cover various fields. NLP activities include creating and identifying text, translating, summarising and rewriting information, classifying and categorising, sentiment analysis, healthcare and conversational AI, and chatbots. The ease of the training process, accuracy, performance, flexibility, extensibility, and adaptability are the advantages that LLMs offer to researchers, general users, and organisations [348]. The data used for inference and training are significantly expanded using an LLM, greatly enhancing the AI model's capabilities. An LLM must be trained on a vast volume of data at the base layer. The data corpus includes at least one billion parameters, typically petabytes in size [349]. The training process starts with unsupervised learning and proceeds in several stages. The method uses unstructured and unlabelled data to train the model. Since unlabelled data are readily available, it results in a smooth training process. The model elucidates connections between various related terms and contexts [350]. Self-supervised learning is the next phase for some LLMs, which involves training and fine-tuning. Labelled data help the model identify different concepts more precisely [351]. As the LLM completes the transformer NN process, DL is applied. The transformer model architecture enables the LLM to comprehend and recognise the connections between content and related concepts through a self-attention mechanism. The system assigns a particular item (referred to as a token) a score, also known as a weight. Once an LLM has been trained, a foundation is established, and AI can be effectively applied [352]. The AI model inference can respond by posing a prompt to the LLM, and the response could take the form of newly created text, summarised text, or a sentiment analysis report. LLMs will be trained continuously on larger datasets, and as fact-checking tools are added, the data are reviewed for correctness and potential bias. Their capacity to adapt material to various contexts will develop further, making them more likely to be used for technical competence. Future LLMs will show better performance and present attributing results. There are numerous limitations to LLMs as well, including development and operating costs, AI hallucination, bias, complexity, explainability, and glitch tokens. The prevalence of misleading information on the internet and in-text corpora puts the dependability and security of LLMs in danger [353]. It highlights the urgent need to comprehend the mechanisms that affect the actions of LLMs due to incorrect information. Data pollution caused by using LLMs presents another major challenge. Moreover, LLMs can generate FN and misleading information, which may flood OSNs and online sources [354]. These challenges of LLMs presented in this review provide prospective research directions to the research community.

- **ChatGPT 3.5 and 4.0:** The Open-AI-trained LLM uses GPT technology and offers generative replies to queries. Its knowledge cutoff restricts the inclusion of web material after a specific date in its training, which is a distinctive feature. The ability to comprehend and produce text is where GPT-3.5 and GPT-4 differ. GPT-4, a more recent edition, is more effective, has been trained on a larger dataset, and has increased performance in producing cohesive and contextually relevant texts [339]. The large number of parameters enables it to understand more intricate linguistic patterns. Additionally, GPT-4 performs better for advanced understanding, text comprehension, and abstraction. Due to a comparable knowledge cutoff, the models cannot respond to information or events of real-world scenarios after their respective training durations [355].
- **Large Language Models AI (LLaMA):** The foundation language models presented by LLaMA range in size from 7 B to 65 B parameters, developed by a group of Meta researchers. They adopted a novel strategy by training models on trillions of freely

accessible tokens for a larger audience. The LLaMA developers focused on scaling the model performance and increased the training data volume rather than the number of parameters. It reduced the computational training cost of LLaMA. LLaMA 2 uses only publicly accessible datasets to deliver high performance. The developers reported that the 13 B parameter model outperformed the considerably bigger GPT-3 (with 175 B parameters) on a wide range of NLP benchmarks, and the largest model was scaled with cutting-edge models, PaLM and Chinchilla. There are slight architectural variations. Contrary to GPT-3, LLaMA employs root-mean-squared layer normalisation, SwiGLU activation function, and rotary positional embeddings, and increases the context length from 2 K to 4 K tokens (between LLaMA 1 and 2) [352].

- **Language Model for Dialogue Applications (LaMDA):** Google's LLM for conversation applications is LaMDA/Bard. Contrary to ChatGPT, it addresses the challenges that users encounter in accessing the internet to produce responses [338]. LaMDA employs a transformer language model that only has decoders. It is trained with fine-tuning data produced by human-annotated safety, interestingness, and responses for sensibleness after being pre-trained on a text corpus consisting of documents and dialogues totalling 1.56 trillion words [356]. Google's tests revealed that LaMDA outperformed human responses in terms of interestingness. The LaMDA transformer model and an external information retrieval system interact to improve the accuracy of facts provided to the user [356].
- **Bing AI or Sydney:** Bing AI, also known as Sydney, is Microsoft's LLM product, and it is based on the Prometheus AI model. It differs from ChatGPT, in that it does not restrict access to online data [338].

6. Results and Discussion

The practical applicability of this review results is multidimensional. The evaluation of interdisciplinary theories, datasets, Fake News Detection tools and websites, and techniques presents an opportunity for future research directions to present an automated mechanism for FND in different contexts. Firstly, the analysis of interdisciplinary theories on fake news (FN) determines how various FN features are craftily designed to instigate the readers and target their associations. This persuades and motivates them to share unverified FN intentionally on the OSNs. Moreover, these theories identify several characteristics, such as the writing style, sources, quality, comments, etc., of the FN that can enlighten users about its veracity. Likewise, theories on OSN users focused on users' behaviours and intentions reveal the motivation behind intentional and unintentional FN sharing on OSNs. The in-depth analysis of these theories shows that the group polarisation and psychological tension caused by the targeted FN or trending stories lead to the unintentional distribution of FN on OSNs. Therefore, these theories provide insight into fundamental factors which should be considered when designing automated mechanisms for FND.

Secondly, the analysis of publicly available datasets determines the need for verifiable, large-sized and multilingual datasets. The lack of multilingual datasets and FND mechanisms in different languages has created a void and language bias in the research. Thus, the significance of such datasets remains pivotal since the automated FND mechanism relies on these datasets. There are around 7000 different languages spoken in the world with different cultural contexts. Therefore, the development and application of multilingual and culturally adaptive automated FND mechanisms in real-world scenarios are crucial to tackling disinformation on OSNs. The related complexities of such FND systems can be overcome by considering the corresponding interdisciplinary theories, which present a potential area of research for future studies.

Thirdly, the evaluation of fact-checking tools and websites necessitates automated FND mechanisms. The limited availability of fact-checked news, especially in resource-constrained languages, and news from limited domains and regions are the limitations of these fact-checking services. This strengthens the case for an effective automated mechanism which can be deployed on a larger scale in real-world scenarios.

Lastly, an analysis of various ML and DL techniques discloses their advantages and disadvantages. This further clarifies that the strengths of Large Language Models (LLMs) should be harnessed for developing large-scale, multilingual FND mechanisms focused on the context of the shared news items. Moreover, these models can be hybrids, where their application can enhance the accuracy and reliability of the shared content with language and cultural context without any biases that may affect the legitimacy of such automated mechanisms.

7. Challenges and Potential Future Directions

The review has reviewed and analysed the existing literature in depth. It concludes that automated FND remains ambiguous despite extensive research and development. There exists much space for further improvement in this field. We discuss the emerging trends for further investigation in FND research and outline potential research areas in the future.

- With the emergence of DL-based and AI generative LLMs, comprehensive datasets that contain cross-language, multilingual (including resource-constrained languages), multimodal, cross-domain and cross-topic analysis are attainable. Moreover, multilingual unlabelled datasets must be generated, as this area has received less attention in the literature. The curation of such datasets may provide large-sized data repositories required to evaluate the limits of the models. It will further enable automated real-time FND, which is needed for misleading content and FN on OSNs and online websites.
- The news stories, clickbait, misleading content, and FN with manipulated and biased facts employ multimedia content (images or videos) to entice online readers for widespread proliferation. The existing literature is focused on content-based analysis significantly and disregards this visual-features investigation. This area remains under-researched for several low-resource languages due to a lack of reasonably sized datasets. It offers an open research opportunity to the research community for FND. Another prospective research area involves creating models that can be implemented for FND on OSNs by combining textual and visual features with correlation analysis of these features. Developing news feed algorithms resistant to FN dissemination has not established sufficient consideration. It is necessary to dismiss echo chambers and biased search engines to detect and eliminate FN.
- Limited research studies concentrate on the purpose of misleading information, while extensive studies use authenticity as a criterion to assess FN. One technique to understand the motivation behind a news item is through expert data annotation. Another way is to choose features carefully and consider recognised social science ideas. The subjects of brain research, neuropsychology, psychology, and other transdisciplinary expertise are comparatively cutting edge. Limited research has been conducted on the neurological mechanisms underlying FN. We are convinced that the unique tools available in this area can help with FND, defence, and comprehension.
- The limits of pre-trained transformer models have not been fully evaluated for NLP-related tasks. Ensemble techniques and multimodel-based approaches using the pre-trained transformer models must be evaluated to extract the hidden features of FN. An upsurge of LLMs has emerged with the development of ChatGPT, etc. In terms of language proficiency, these models outperform the existing models. There is potential to enhance FND performance by exploiting the knowledge and linguistic skills of the models. However, there is still a lack of studies on large-scale FND.
- Deepfake technology is becoming more pervasive in different fields, such as games, film and television, privacy protection, and research (in terms of plagiarised material) in the age of AI-generated content. Deepfake technology poses a threat to users' reputations, social order, and political security when used maliciously. Therefore, future research should concentrate on creating deep forgeries creation and protection techniques to overcome the aforementioned challenges.

- Another potential direction for LLMs in the future is to enable more precise information through domain-specific LLMs created for specific fields and industries. The accuracy and performance of LLMs can be enhanced with a large-scale application of methods, i.e., reinforcement learning from human feedback. There is a need to create comprehensive and comprehensible solutions for explainable FND. Explainable AI techniques for FND are still in their nascent stages of development, both in terms of model structure analysis and model behaviour analysis. Developing and implementing an FND system in the current scenario for AI-generated information distribution context is essential.
- Automated FND may lead to automated censorship on OSNs and other online web portals. Therefore, it is significant to strike a balance between automated FND and automated censorship. Since the automated FND mechanisms depend on the algorithms, transparency and comprehensiveness of the decision-making process of these algorithms without any bias could mitigate the concerns about undue censorship. Another applicable approach could be a hybrid system, where the right of the final decision could be with human experts. The final decisions on inclusion and exclusion should also have clear criteria and guidelines. Lastly, policy frameworks, regulations, and full-fledged auditing can strike a balance between automated FND and automated censorship.

8. Conclusions

The volume and velocity of FN distribution on online websites, blogs, and OSNs is alarming. FN confounds its readers, creates distrust in media, and influences national and international spectra in various fields. The comprehensive analysis of existing FND approaches and techniques have inferred that the literature provides limited automated insights for FND. The proposed methods and techniques in the existing literature undermine the effectiveness of interdisciplinary theories on FN and OSN users. These theories highlight the incitement of intentional and unintentional FND propagation. Thus, designing the FND systems in light of the proposed recommendations that expose FN-related biases and motives is significant. The constant development of publicly available datasets is remarkable. However, the FND datasets analysis identified the limited number of large-sized, multidomain, multiclass, multilingual (including low-resource languages), mixed-lingual (high- and low-resource languages), labelled, and unlabelled datasets. Future research with benchmark datasets should concentrate their efforts in the identified directions. It will provide a large depository of datasets to investigate the efficacy of pre-trained transformer models and LLMs for automated FND on OSNs and online sources in different languages and domains. Thirdly, NLP, pre-trained models, and SOTA LLMs have proved their efficacy and capability for various language-related tasks. Therefore, deepfakes and AI-generated content creation and widespread dissemination can be confined using the SOTA techniques. LLMs offer an opportunity for researchers to investigate, develop, and implement effective methods to fight fire with fire, which may provide an automated FND mechanism.

Author Contributions: Conceptualization, M.A.A.; Validation, S.H.; Resources, N.A.; Data curation, H.J.H.; Writing—original draft, S.H. and M.A.A.; Funding acquisition, M.A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Prince Sultan University for paying the Article Processing Charges (APC) of this publication.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
ALBERT	A Lite BERT
ANN	Artificial Neural Network
BERT	Bidirectional Encoder Representations from Transformers
BoW	Bag of Words
BSMM	Boundary Samples Mining Module
BRNNs	Bidirectional Recurrent Neural Networks
CLIP	Contrastive Language-Image Pre-Training
CNN	Convolutional Neural Network
DBM	Deep Boltzmann Machine
DeBERTa	Decoding-enhanced BERT with Disentangled Attention
DL	Deep Learning
DNNs	Deep Neural Networks
DT	Decision Tree
ELECTRA	Efficiently Learning an Encoder that Classifies Token Replacements Accurately
FN	Fake News
FND	Fake News Detection
FSM	Future Scaling Module
GA	Genetic Algorithm
GAN	Generative Adversarial Network
GDL	Graphical Deep Learning
GLUE	General Language Understanding Evaluation
GPT	Generative Pre-trained Transformer
GRU	Gated Recurrent Unit
LaMDA	Language Model for Dialogue Applications
LIWC	Linguistic Inquiry and Word Count
LLaMA	Large Language Model Meta AI
LLMs	Large Language Models
LR	Logistic Regression
LSTM	Long Short-Term Memory
MARGE	Multilingual Autoencoder that Retrieves and Generates
mBART	Multilingual BART
mBERT	Multilingual BERT
MCC	Matthew's Correlation Coefficient
ML	Machine Learning
MLM	Masked Language Model
MLP	Multilayer Perceptron
NLP	Natural Language Processing
NSP	Next Sentence Prediction
OSNs	Online Social Networks
PaLM	Pathways Language Model
PLM	Permutated Language Modelling
RBM	Restricted Boltzmann Machine
ResNet	Residual Network
RF	Random Forest
RNN	Recurrent Neural Network
RoBERTa	Robustly optimised BERT Approach
SAHS	Self-Adaptive Harmony Search
SBUs	Stochastic Binary Units
SOP	Sentence Order Prediction
SOTA	State of the Art
SQuAD	Stanford Question Answering Dataset
SVM	Support Vector Machine
SWAG	Situations With Adversarial Generations
SwiGLU	Swish Gated Linear Unit

Swin	Shifted Window
TF-IDF	Term Frequency-Inverse Document Frequency
TL	Transfer Learning
VGG	Visual Geometry Group

References

1. Nguyen, V.H.; Sugiyama, K.; Nakov, P.; Kan, M.Y. Fang: Leveraging social context for fake news detection using graph representation. In Proceedings of the 29th ACM International Conference on information & Knowledge Management, Seoul, Republic of Korea, 19 October 2020.
2. Bondielli, A.; Marcelloni, F. A review on fake news and rumour detection techniques. *Inf. Sci.* **2019**, *497*, 38–55. [CrossRef]
3. Staff, M.M.S. Understanding the Fake News Universe, Media Matters for America. 2016. Available online: https://www.mediamatters.org/fake-news/understanding-fake-news-universe (accessed on 20 April 2023).
4. Rubin, V.L. On deception and deception detection: Content analysis of computer-mediated stated beliefs. *Proc. Am. Soc. Inf. Sci. Technol.* **2010**, *47*, 1–10. [CrossRef]
5. Park, M.; Chai, S. Constructing a User-Centered Fake News Detection Model by Using Classification Algorithms in Machine Learning Techniques. *IEEE Access* **2023**, *11*, 71517–71527. [CrossRef]
6. Chang, Y.; Wang, X. Detecting fake news via deep learning techniques. In Proceedings of the ICMLCA 2021 2nd International Conference on Machine Learning and Computer Application, Shenyang, China, 17–19 December 2021; VDE: Hannover, Germany, 2021; pp. 1–4.
7. Islam, M.R.; Liu, S.; Wang, X.; Xu, G. Deep learning for misinformation detection on online social networks: A review and new perspectives. *Soc. Netw. Anal. Min.* **2020**, *10*, 82. [CrossRef] [PubMed]
8. Sengupta, E.; Nagpal, R.; Mehrotra, D.; Srivastava, G. ProBlock: A novel approach for fake news detection. *Clust. Comput.* **2021**, *24*, 3779–3795. [CrossRef] [PubMed]
9. Ashcroft, M.; Fisher, A.; Kaati, L.; Omer, E.; Prucha, N. Detecting jihadist messages on twitter. In Proceedings of the 2015 European Intelligence and Security Informatics Conference, Manchester, UK, 7–9 September 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 161–164.
10. Konkobo, P.M.; Zhang, R.; Huang, S.; Minoungou, T.T.; Ouedraogo, J.A.; Li, L. A deep learning model for early detection of fake news on social media. In Proceedings of the 7th International Conference on Behavioural and Social Computing, Bournemouth, UK, 5–7 November 2020; pp. 1–6.
11. Zafarani, R.; Zhou, X.; Shu, K.; Liu, H. Fake news research: Theories, detection strategies, and open problems. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, New York, NY, USA, 25 July 2019; pp. 3207–3208.
12. Zhou, X.; Zafarani, R. A review of fake news: Fundamental theories, detection methods, and opportunities. *ACM Comput. Rev. (CSUR)* **2020**, *53*, 1–40.
13. Lim, C. Checking how fact-checkers check. *Res. Politics* **2018**, *5*, 2053168018786848. [CrossRef]
14. Nickerson, R.S. Confirmation bias: A ubiquitous phenomenon in many guises. *Rev. Gen. Psychol.* **1998**, *2*, 175–220. [CrossRef]
15. Freedman, J.L.; Sears, D.O. Selective exposure. In *Advances in Experimental Social Psychology*; Academic Press: Cambridge, MA, USA, 1965; Volume 2, pp. 57–97.
16. Metzger, M.J.; Hartsell, E.H.; Flanagin, A.J. Cognitive dissonance or credibility? A comparison of two theoretical explanations for selective exposure to partisan news. *Commun. Res.* **2020**, *47*, 3–28. [CrossRef]
17. Boehm, L.E. The validity effect: A search for mediating variables. *Personal. Soc. Psychol. Bull.* **1994**, *20*, 285–293. [CrossRef]
18. Zhou, X.; Zafarani, R.; Shu, K.; Liu, H. Fake news: Fundamental theories, detection strategies and challenges. In Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, New York, NY, USA, 30 January 2019; pp. 836–837.
19. Leibenstein, H. Bandwagon, snob, and Veblen effects in the theory of consumers' demand. *Q. J. Econ.* **1950**, *64*, 183–207. [CrossRef]
20. Kim, G.; Ko, Y. Effective fake news detection using graph and summarization techniques. *Pattern Recognit. Lett.* **2021**, *151*, 135–139. [CrossRef]
21. Habib, A.; Asghar, M.Z.; Khan, A.; Habib, A.; Khan, A. False information detection in online content and its role in decision making: A systematic literature review. *Soc. Netw. Anal. Min.* **2019**, *9*, 50. [CrossRef]
22. Singh, B.; Sharma, D.K. Predicting image credibility in fake news over social media using multi-modal approach. *Neural Comput. Appl.* **2022**, *34*, 21503–21517. [CrossRef] [PubMed]
23. Sharma, D.K.; Garg, S. IFND: A benchmark dataset for fake news detection. *Complex Intell. Syst.* **2021**, *9*, 2843–2863. [CrossRef]
24. Chen, Y. Convolutional Neural Network for Sentence Classification. Master's Thesis, University of Waterloo, Waterloo, ON, Canada, 2015.
25. Oukali, S.; Lazri, M.; Labadi, K.; Brucker, J.M.; Ameur, S. Development of a hybrid classification technique based on deep learning applied to MSG/SEVIRI multispectral data. *J. Atmos. Sol.-Terr. Phys.* **2019**, *193*, 105062. [CrossRef]
26. Olteanu, A.; Castillo, C.; Diaz, F.; Kıcıman, E. Social Data: Biases, methodological pitfalls, and ethical boundaries. *Front. Big Data* **2019**, *2*, 13. [CrossRef]
27. Lazer, D.M.; 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]

28. De Beer, D.; Matthee, M. Approaches to identify fake news: A systematic literature review. In Proceedings of the Integrated Science in Digital Age 2020, ICIS 2020, Kep, Cambodia, 1–3 May 2020; Springer: Cham, Switzerland, 2021; pp. 13–22.
29. Silva, A.; Han, Y.; Luo, L.; Karunasekera, S.; Leckie, C. Propagation2Vec: Embedding partial propagation networks for explainable fake news early detection. *Inf. Process. Manag.* **2021**, *58*, 102618. [CrossRef]
30. Dabbous, A.; Aoun Barakat, K.; de Quero Navarro, B. Fake news detection and social media trust: A cross-cultural perspective. *Behav. Inf. Technol.* **2022**, *41*, 2953–2972. [CrossRef]
31. Meneses Silva, C.V.; Silva Fontes, R.; Colaço Júnior, M. Intelligent fake news detection: A systematic mapping. *J. Appl. Secur. Res.* **2021**, *16*, 168–189. [CrossRef]
32. Krishna, S.R.; Vasantha, S.V.; Deep, K.M. Review on fake news detection using machine learning algorithms. *Int. J. Eng. Res. Technol. (IJERT)* **2021**, *9*, 121–125.
33. Ribeiro Bezerra, J.F. Content-based fake news classification through modified voting ensemble. *J. Inf. Telecommun.* **2021**, *5*, 499–513. [CrossRef]
34. Vereshchaka, A.; Cosimini, S.; Dong, W. Analyzing and distinguishing fake and real news to mitigate the problem of disinformation. *Comput. Math. Organ. Theory* **2020**, *26*, 350–364. [CrossRef]
35. Alghamdi, J.; Lin, Y.; Luo, S. Towards COVID-19 fake news detection using transformer-based models. *Knowl.-Based Syst.* **2023**, *274*, 110642. [CrossRef]
36. Shaik, M.A.; Sree, M.Y.; Vyshnavi, S.S.; Ganesh, T.; Sushmitha, D.; Shreya, N. Fake News Detection using NLP. In Proceedings of the 2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA), Uttarakhand, India, 14–16 March 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 399–405.
37. Ibrishimova, M.D.; Li, K.F. A machine learning approach to fake news detection using knowledge verification and natural language processing. In *Advances in Intelligent Networking and Collaborative Systems, Proceedings of the 11th International Conference on Intelligent Networking and Collaborative Systems (INCoS-2019), Oita, Japan, 5–7 September 2019*; Springer International Publishing: Cham, Switserland, 2020; pp. 223–234.
38. Kula, S.; Choraś, M.; Kozik, R.; Ksieniewicz, P.; Woźniak, M. Sentiment analysis for fake news detection by means of neural networks. In Proceedings of the Computational Science—ICCS 2020: 20th International Conference 2020, Amsterdam, The Netherlands, 3–5 June 2020; Springer International Publishing: Cham, Swizterland, 2020; Proceedings, Part IV 20; pp. 653–666.
39. Popat, K. Assessing the credibility of claims on the web. In Proceedings of the 26th International Conference on World Wide Web Companion, Geneva, Switzerland, 3–7 April 2017; pp. 735–739.
40. Sampson, J.; Morstatter, F.; Wu, L.; Liu, H. Leveraging the implicit structure within social media for emergent rumor detection. In Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, New York, NY, USA, 24–28 October 2016; pp. 2377–2382.
41. Wu, K.; Yang, S.; Zhu, K.Q. False rumors detection on sina weibo by propagation structures. In Proceedings of the 31st International Conference on Data Engineering, Seoul, Republic of Korea, 13–17 April 2015; pp. 651–662.
42. Cooke, N.A. Posttruth, truthiness, and alternative facts: Information behavior and critical information consumption for a new age. *Libr. Q.* **2017**, *87*, 211–221. [CrossRef]
43. 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.* **2021**, *65*, 180–212. [CrossRef]
44. Grieve, J.; Woodfield, H. *The Language of Fake News*; Cambridge University Press: Cambridge, UK, 2023.
45. Dame Adjin-Tettey, T. Combating fake news, disinformation, and misinformation: Experimental evidence for media literacy education. *Cogent Arts Humanit.* **2022**, *9*, 2037229. [CrossRef]
46. Shu, K.; Wang, S.; Lee, D.; Liu, H. Mining disinformation and fake news: Concepts, methods, and recent advancements. In *Disinformation, Misinformation, and Fake News in Social Media: Emerging Research Challenges and Opportunities*; Springer: Cham, Swizterland, 2020; pp. 1–19.
47. Kapantai, E.; Christopoulou, A.; Berberidis, C.; Peristeras, V. A systematic literature review on disinformation: Toward a unified taxonomical framework. *New Media Soc.* **2021**, *23*, 1301–1326. [CrossRef]
48. Meel, P.; Vishwakarma, D.K. Fake news, rumor, information pollution in social media and web: A contemporary review of state-of-the-arts, challenges and opportunities. *Expert Syst. Appl.* **2020**, *153*, 112986. [CrossRef]
49. Monsees, L. Information disorder, fake news and the future of democracy. *Globalizations* **2023**, *20*, 153–168. [CrossRef]
50. Guadagno, R.E.; Guttieri, K. Fake news and information warfare: An examination of the political and psychological processes from the digital sphere to the real world. In *Research Anthology on Fake News, Political Warfare, and Combatting the Spread of Misinformation*; IGI Global: Pennsylvania, PA, USA, 2021; pp. 218–242.
51. Dowse, A.; Bachmann, S.D. Information warfare: Methods to counter disinformation. *Def. Secur. Anal.* **2022**, *38*, 450–469. [CrossRef]
52. We Are Social. The Changing World of Digital in 2023. Available online: https://wearesocial.com/us/blog/2023/01/the-changing-world-of-digital-in-2023/ (accessed on 20 April 2024).
53. UTK Machine Learning Club. Fake News: Build a System to Identify Unreliable News Articles. 2018. Available online: https://www.kaggle.com/c/fake-news (accessed on 20 April 2024).
54. Tandoc, E.C.; Lee, J.; Chew, M.; Tan, F.X.; Goh, Z.H. Falling for fake news: The role of political bias and cognitive ability. *Asian J. Commun.* **2021**, *31*, 237–253. [CrossRef]

55. DeMers, J. 59 percent of you will share this article without even reading it. *Forbes*, 20 April 2020.
56. Metzger, M.J.; Flanagin, A.J. Using Web 2.0 technologies to enhance evidence-based medical information. *J. Health Commun.* **2011**, *16* (Suppl. S1), 45–58. [CrossRef] [PubMed]
57. O'Connell, E. Navigating the internet's information cesspool, fake news and what to do about it. *Univ. Pac. Law Rev.* **2022**, *53*, 251.
58. Thorson, E. Belief echoes: The persistent effects of corrected misinformation. *Political Commun.* **2016**, *33*, 460–480. [CrossRef]
59. Metzger, M.J.; Flanagin, A.J.; Mena, P.; Jiang, S.; Wilson, C. From dark to light: The many shades of sharing misinformation online. *Media Commun.* **2021**, *9*, 134–143. [CrossRef]
60. Newman, N.; Fletcher, R.; Schulz, A.; Andi, S.; Robertson, C.T.; Nielsen, R.K. *Reuters Institute Digital News Report 2021*; Reuters Institute for the Study of Journalism: Oxford, UK, 2021.
61. Saeed, M.; Traub, N.; Nicolas, M.; Demartini, G.; Papotti, P. Crowdsourced fact-checking at Twitter: How does the crowd compare with experts? In Proceedings of the 31st ACM International Conference on Information & Knowledge Management, Atlanta, GA, USA, 17–21 October 2022; pp. 1736–1746.
62. Martens, B.; Aguiar, L.; Gomez-Herrera, E.; Mueller-Langer, F. *The Digital Transformation of News Media and the Rise of Disinformation and Fake News*; European Commission: Seville, Spain, 2018.
63. Strömbäck, J.; Tsfati, Y.; Boomgaarden, H.; Damstra, A.; Lindgren, E.; Vliegenthart, R.; Lindholm, T. News media trust and its impact on media use: Toward a framework for future research. *Ann. Int. Commun. Assoc.* **2020**, *44*, 139–156. [CrossRef]
64. Mitra, T.; Gilbert, E. Credbank: A large-scale social media corpus with associated credibility annotations. In Proceedings of the International AAAI Conference on Web and Social Media, Buffalo, NY, USA, 2–6 June 2015; Volume 9, No. 1, pp. 258–267.
65. Zhou, X.; Zafarani, R. Fake news: A review of research, detection methods, and opportunities. *arXiv* **2018**, arXiv:1812.00315.
66. Shu, K.; Sliva, A.; Wang, S.; Tang, J.; Liu, H. Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explor. Newsl.* **2017**, *19*, 22–36. [CrossRef]
67. Yuan, L.; Jiang, H.; Shen, H.; Shi, L.; Cheng, N. Sustainable development of information dissemination: A review of current fake news detection research and practice. *Systems* **2023**, *11*, 458. [CrossRef]
68. Sharma, K.; Qian, F.; Jiang, H.; Ruchansky, N.; Zhang, M.; Liu, Y. Combating fake news: A review on identification and mitigation techniques. *ACM Trans. Intell. Syst. Technol. (TIST)* **2019**, *10*, 1–42. [CrossRef]
69. Lao, A.; Shi, C.; Yang, Y. Rumor detection with field of linear and non-linear propagation. In Proceedings of the International World Wide Web Conferences, New York, NY, USA, 19–23 April 2021; pp. 3178–3187.
70. Zhang, X.; Ghorbani, A.A. An overview of online fake news: Characterization, detection, and discussion. *Inf. Process. Manag.* **2020**, *57*, 102025. [CrossRef]
71. de Oliveira, N.R.; Pisa, P.S.; Lopez, M.A.; de Medeiros, D.S.V.; Mattos, D.M. Identifying fake news on social networks based on natural language processing: Trends and challenges. *Information* **2021**, *12*, 38. [CrossRef]
72. Kim, B.; Xiong, A.; Lee, D.; Han, K. A systematic review on fake news research through the lens of news creation and consumption: Research efforts, challenges, and future directions. *PLoS ONE* **2021**, *16*, e0260080. [CrossRef]
73. 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]
74. Mridha, M.F.; Keya, A.J.; Hamid, M.A.; Monowar, M.M.; Rahman, M.S. A comprehensive review on fake news detection with deep learning. *IEEE Access* **2021**, *9*, 156151–156170. [CrossRef]
75. D'Ulizia, A.; Caschera, M.C.; Ferri, F.; Grifoni, P. Fake news detection: A review of evaluation datasets. *PeerJ Comput. Sci.* **2021**, *7*, e518. [CrossRef]
76. Murayama, T. Dataset of fake news detection and fact verification: A review. *arXiv* **2021**, arXiv:2111.03299.
77. Collins, B.; Hoang, D.T.; Nguyen, N.T.; Hwang, D. Trends in combating fake news on social media—A review. *J. Inf. Telecommun.* **2021**, *5*, 247–266.
78. Khan, T.; Michalas, A.; Akhunzada, A. Fake news outbreak 2021: Can we stop the viral spread? *J. Netw. Comput. Appl.* **2021**, *190*, 103112. [CrossRef]
79. Rohera, D.; Shethna, H.; Patel, K.; Thakker, U.; Tanwar, S.; Gupta, R.; Hong, W.C.; Sharma, R. A taxonomy of fake news classification techniques: Review and implementation aspects. *IEEE Access* **2022**, *10*, 30367–30394. [CrossRef]
80. Hu, L.; Wei, S.; Zhao, Z.; Wu, B. Deep learning for fake news detection: A comprehensive review. *AI Open* **2022**, *3*, 133–155. [CrossRef]
81. Hangloo, S.; Arora, B. Combating multimodal fake news on social media: Methods, datasets, and future perspective. *Multimed. Syst.* **2022**, *28*, 2391–2422. [CrossRef] [PubMed]
82. Rastogi, S.; Bansal, D. A review on fake news detection 3T's: Typology, time of detection, taxonomies. *Int. J. Inf. Secur.* **2023**, *22*, 177–212. [CrossRef]
83. Ruffo, G.; Semeraro, A.; Giachanou, A.; Rosso, P. Studying fake news spreading, polarisation dynamics, and manipulation by bots: A tale of networks and language. *Comput. Sci. Rev.* **2023**, *47*, 100531. [CrossRef]
84. Kondamudi, M.R.; Sahoo, S.R.; Chouhan, L.; Yadav, N. A comprehensive review of fake news in social networks: Attributes, features, and detection approaches. *J. King Saud Univ.–Comput. Inf. Sci.* **2023**, *35*, 101571.
85. Tranfield, D.; Denyer, D.; Smart, P. Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *Br. J. Manag.* **2003**, *14*, 207–222. [CrossRef]

86. Allen, V.L.; Wilder, D.A.; Atkinson, M.L. Multiple group membership and social identity. In *Studies in Social Identity*; Sarbin, T.R., Scheibe, K.E., Eds.; Praeger: New York, NY, USA, 1983; pp. 92–115.
87. Ashforth, B.E.; Mael, F. Social identity theory and the organization. *Acad. Manag. Rev.* **1989**, *14*, 20–39. [CrossRef]
88. Zhou, X.; Jain, A.; Phoha, V.V.; Zafarani, R. Fake news early detection: A theory-driven model. *Digit. Threat. Res. Pract.* **2020**, *1*, 1–25. [CrossRef]
89. Silva, A.; Luo, L.; Karunasekera, S.; Leckie, C. Embracing domain differences in fake news: Cross-domain fake news detection using multi-modal data. In Proceedings of the AAAI Conference on Artificial Intelligence, Melbourne, Australia, 18 May 2021; Volume 35, pp. 557–565.
90. Heidegger, M. *Being and Time*; Suny Press: Albany, NY, USA, 2010.
91. Undeutsch, U. Beurteilung der glaubhaftigkeit von aussagen. *Handb. Der Psychol.* **1967**, *11* (Suppl. S126), 26–181.
92. Burke, K. *A Rhetoric of Motives*; University of California Press: Berkeley, CA, USA, 1969.
93. Johnson, M.K.; Raye, C.L. Reality monitoring. *Psychol. Rev.* **1981**, *88*, 67. [CrossRef]
94. Erfurth, W. Surprise. In *The Roots of Strategy, Book 3. Military Classics*; Possony, S., Vilfroy, D., Eds.; Stagpole Books: Harrisburg, PA, USA, 1943; pp. 385–547.
95. Handel, M.I. *Introduction: Strategic and Operational Deception in Historical Perspective*; Taylor and Francis: London, UK, 1987.
96. Zuckerman, M.; DePaulo, B.M.; Rosenthal, R. Verbal and nonverbal communication of deception. In *Advances in Experimental Social Psychology*; Academic Press: Cambridge, MA, USA, 1981; Volume 14, pp. 1–59.
97. Wittgenstein, L. *Zettel*; University of California Press: Berkeley, CA, USA, 1967.
98. McCornack, S.A.; Morrison, K.; Paik, J.E.; Wisner, A.M.; Zhu, X. Information manipulation theory 2: A propositional theory of deceptive discourse production. *J. Lang. Soc. Psychol.* **2014**, *33*, 348–377. [CrossRef]
99. Flanagin, A.; Metzger, M.J. Digital Media and Perceptions of Source Credibility in Political Communication. In *The Oxford Handbook of Political Communication*; Oxford Academic: Oxford, UK, 2017; p. 417.
100. Deutsch, M.; Gerard, H.B. A study of normative and informational social influences upon individual judgment. *J. Abnorm. Soc. Psychol.* **1955**, *51*, 629. [CrossRef]
101. Compton, J. Inoculation theory. In *The SAGE Handbook of Persuasion: Developments in Theory and Practice*; SAGE Publications, Inc.: Thousand Oaks, CA, USA, 2013; Volume 2, pp. 220–237.
102. Kuran, T.; Sunstein, C.R. Availability cascades and risk regulation. *Stan. L. Rev.* **1998**, *51*, 683. [CrossRef]
103. Gwebu, K.L.; Wang, J.; Zifla, E. Can warnings curb the spread of fake news? The interplay between warning, trust and confirmation bias. *Behav. Inf. Technol.* **2022**, *41*, 3552–3573. [CrossRef]
104. Basu, S. The conservatism principle and the asymmetric timeliness of earnings1. *J. Account. Econ.* **1997**, *24*, 3–37. [CrossRef]
105. Baumann, F.; Lorenz-Spreen, P.; Sokolov, I.M.; Starnini, M. Modeling echo chambers and polarization dynamics in social networks. *Phys. Rev. Lett.* **2020**, *124*, 048301. [CrossRef]
106. Fisher, R.J. Social desirability bias and the validity of indirect questioning. *J. Consum. Res.* **1993**, *20*, 303–315. [CrossRef]
107. Bálint, P.; Bálint, G. The semmelweis-reflex. *Orvosi Hetil.* **2009**, *150*, 1430. [CrossRef]
108. Ross, L.; Ward, A. Naive realism in everyday life: Implications for social conflict and misunderstanding. In *Values and Knowledge*; Psychology Press: London, UK, 2013.
109. MacLeod, C.; Mathews, A.; Tata, P. Attentional bias in emotional disorders. *J. Abnorm. Psychol.* **1986**, *95*, 15. [CrossRef] [PubMed]
110. Bruce, D.; Papay, J.P. Primacy effect in single-trial free recall. *J. Verbal Learn. Verbal Behavior.* **1970**, *9*, 473–486. [CrossRef]
111. Festinger, L. *A Theory of Cognitive Dissonance*; Stanford University Press: Redwood City, CA, USA, 1957.
112. Levy, J.S. An introduction to prospect theory. *Political Psychol.* **1992**, *13*, 171–186.
113. Dunning, D.; Griffin, D.W.; Milojkovic, J.D.; Ross, L. The overconfidence effect in social prediction. *J. Personal. Soc. Psychol.* **1990**, *58*, 568. [CrossRef] [PubMed]
114. Frijda, N.H. *The Emotions*; Cambridge University Press: Cambridge, UK, 1986.
115. Pronin, E.; Kruger, J.; Savtisky, K.; Ross, L. You don't know me, but I know you: The illusion of asymmetric insight. *J. Personal. Soc. Psychol.* **2001**, *81*, 639. [CrossRef]
116. Hovland, C.I.; Harvey, O.J.; Sherif, M. Assimilation and contrast effects in reactions to communication and attitude change. *J. Abnorm. Soc. Psychol.* **1957**, *55*, 244. [CrossRef]
117. Ahmed, S. Disinformation sharing thrives with fear of missing out among low cognitive news users: A cross-national examination of intentional sharing of deep fakes. *J. Broadcast. Electron. Media* **2022**, *66*, 89–109. [CrossRef]
118. Raza, S.; Ding, C. A Regularized Model to Trade-off between Accuracy and Diversity in a News Recommender System. In Proceedings of the 2020 IEEE International Conference on Big Data (Big Data), Virtual, 10 December 2020; pp. 551–560.
119. Hai Ming, L.; Gang, L.; Hua, H.; Waqas, M. Modeling the influencing factors of electronic word-of-mouth about CSR on social networking sites. *Environ. Sci. Pollut. Res.* **2022**, *29*, 66204–66221. [CrossRef]
120. Giachanou, A.; Ghanem, B.; Ríssola, E.A.; Rosso, P.; Crestani, F.; Oberski, D. The impact of psycholinguistic patterns in discriminating between fake news spreaders and fact checkers. *Data Knowl. Eng.* **2022**, *138*, 101960. [CrossRef]
121. Ferrara, E.; Varol, O.; Davis, C.; Menczer, F.; Flammini, A. The rise of social bots. Commun. *ACM* **2016**, *59*, 96–104. [CrossRef]
122. Jones, M.O. The gulf information war | propaganda, fake news, and fake trends: The weaponization of twitter bots in the gulf crisis. *Int. J. Commun.* **2019**, *13*, 27.

123. Zannettou, S.; Caulfield, T.; Setzer, W.; Sirivianos M.; Stringhini, G.; Blackburn, J. Who let the trolls out? Towards understanding state-sponsored trolls. In Proceedings of the 10th ACM Conference on Web Science (WebSci), Boston, MA, USA, 30 June–3 July 2019; pp. 353–362.
124. Zhao, Z.; Resnick, P.; Mei, Q. Enquiring minds: Early detection of rumors in social media from enquiry posts. In Proceedings of the 24th International Conference on World Wide Web, New York, NY, USA, 18–25 May 2015; pp. 1395–1405.
125. Bakshy, E.; Messing, S.; Adamic, L.A. Exposure to ideologically diverse news and opinion on Facebook. *Science* **2015**, *348*, 1130–1132. [CrossRef] [PubMed]
126. Wang, R.; He, Y.; Xu, J.; Zhang, H. Fake news or bad news? Toward an emotion-driven cognitive dissonance model of misinformation diffusion. *Asian J. Commun.* **2020**, *30*, 317–342. [CrossRef]
127. Harper, L.; Herbst, K.W.; Bagli, D.; Kaefer, M.; Beckers, G.M.A.; Fossum, M.; Kalfa, N. The battle between fake news and science. *J. Pediatr. Urol.* **2020**, *16*, 114–115. [CrossRef] [PubMed]
128. Rúas Araujo, J.; Wihbey, J.P.; Barredo-Ibáñez, D. Beyond Fake News and Fact-Checking: A Special Issue to Understand the Political, Social and Technological Consequences of the Battle against Misinformation and Disinformation. *J. Media* **2022**, *3*, 254–256. [CrossRef]
129. Sears, D.O.; Freedman, J.L. Selective exposure to information: A critical review. *Public Opin. Q.* **1967**, *31*, 194–213. [CrossRef]
130. Grimm, P. *Social Desirability Bias*; Wiley International Encyclopedia of Marketing: Hoboken, NJ, USA, 2010.
131. Ostrowski, W.; Arora, A.; Atanasova, P.; Augenstein, I. Multi-hop fact checking of political claims. *arXiv* **2020**, arXiv:2009.06401.
132. Shu, K.; Mahudeswaran, D.; Wang, S.; Lee, D.; Liu, H. Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. *Big Data* **2020**, *8*, 171–188. [CrossRef]
133. Nakamura, K.; Levy, S.; Wang, W.Y. r/fakeddit: A new multimodal benchmark dataset for fine-grained fake news detection. *arXiv* **2019**, arXiv:1911.03854.
134. Ma, J.; Gao, W.; Wong, K.F. *Detect Rumors in Microblog Posts Using Propagation Structure Via Kernel Learning*; Association for Computational Linguistics: Kerrville, TX, USA, 2017.
135. Setty, V.; Rekve, E. Truth be told: Fake news detection using user reactions on reddit. In Proceedings of the 29th ACM International Conference on Information & Knowledge Management, New York, NY, USA, 19–23 October 2020; pp. 3325–3328.
136. Zubiaga, A.; Liakata, M.; Procter, R.; Wong Sak Hoi, G.; Tolmie, P. Analysing how people orient to and spread rumours in social media by looking at conversational threads. *PLoS ONE* **2016**, *11*, e0150989. [CrossRef]
137. Kochkina, E.; Liakata, M.; Zubiaga, A. All-in-one: Multi-task learning for rumour verification. *arXiv* **2018**, arXiv:1806.03713.
138. Sakketou, F.; Plepi, J.; Cervero, R.; Geiss, H.J.; Rosso, P.; Flek, L. Factoid: A new dataset for identifying misinformation spreaders and political bias. *arXiv* **2022**, arXiv:2205.06181.
139. Giachanou, A.; Zhang, G.; Rosso, P. Multimodal fake news detection with textual, visual and semantic information. In *Text 2020, Speech, and Dialogue, Proceedings of the 23rd International Conference, TSD 2020, Brno, Czech Republic, 8–11 September 2020*; Proceedings 23; Springer International Publishing: Cham, Switzerland 2020; pp. 30–38.
140. Ma, J.; Gao, W.; Mitra, P.; Kwon, S.; Jansen, B.J.; Wong, K.F.; Cha, M. Detecting rumors from microblogs with recurrent neural networks. In Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16), New York, NY, USA, 9–15 July 2016.
141. Tam, N.T.; Weidlich, M.; Zheng, B.; Yin, H.; Hung, N.Q.V.; Stantic, B. From anomaly detection to rumour detection using data streams of social platforms. *Proc. VLDB Endow.* **2019**, *12*, 1016–1029. [CrossRef]
142. Derczynski, L.; Bontcheva, K.; Liakata, M.; Procter, R.; Hoi, G.W.S.; Zubiaga, A. SemEval-2017 Task 8: RumourEval: Determining rumour veracity and support for rumours. *arXiv* **2017**, arXiv:1704.05972.
143. Gorrell, G.; Bontcheva, K.; Derczynski, L.; Kochkina, E.; Liakata, M.; Zubiaga, A. Rumoureval 2019: Determining rumour veracity and support for rumours. *arXiv* **2017**, arXiv:1809.06683.
144. Jiang, S.; Wilson, C. Linguistic signals under misinformation and fact-checking: Evidence from user comments on social media. In *Proceedings of the ACM on Human-Computer Interaction 2018, 2(CSCW)*; Association for Computing Machinery: New York, NY, USA, 2018; pp. 1–23.
145. Tacchini, E.; Ballarin, G.; Della Vedova, M.L.; Moret, S.; De Alfaro, L. Some like it hoax: Automated fake news detection in social networks. *arXiv* **2017**, arXiv:1704.07506.
146. Santia, G.; Williams, J. Buzzface: A news veracity dataset with facebook user commentary and egos. In Proceedings of the International AAAI Conference on Web and Social Media, Palo Alto, CA, USA, 25–28 June 2018; Volume 12, pp. 531–540.
147. Reis, J.C.; Melo, P.; Garimella, K.; Almeida, J.M.; Eckles, D.; Benevenuto, F. A dataset of fact-checked images shared on whatsapp during the brazilian and indian elections. In Proceedings of the International AAAI Conference on Web and Social Media, Virtual, 8–11 June 2019; Volume 14, pp. 903–908.
148. Wang, Y.; Yang, W.; Ma, F.; Xu, J.; Zhong, B.; Deng, Q.; Gao, J. Weak supervision for fake news detection via reinforcement learning. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 516–523.
149. Barbado, R.; Araque, O.; Iglesias, C.A. A framework for fake review detection in online consumer electronics retailers. *Inf. Process. Manag.* **2019**, *56*, 1234–1244. [CrossRef]
150. Chen, E.; Lerman, K.; Ferrara, E. Tracking social media discourse about the COVID-19 pandemic: Development of a public coronavirus twitter data set. *JMIR Public Health Surveill.* **2020**, *6*, e19273. [CrossRef]

151. Cui, L.; Lee, D. Coaid: COVID-19 healthcare misinformation dataset. *arXiv* **2020**, arXiv:2006.00885.
152. Dharawat, A.; Lourentzou, I.; Morales, A.; Zhai, C. Drink Bleach or Do What Now? Covid-HeRA: A Study of Risk-Informed Health Decision Making in the Presence of COVID-19 Misinformation. In Proceedings of the International AAAI Conference on Web and Social Media, Atlanta, GA, USA, 6–9 June 2022; Volume 16, pp. 1218–1227.
153. Dai, E.; Sun, Y.; Wang, S. Ginger cannot cure cancer: Battling fake health news with a comprehensive data repository. In Proceedings of the International AAAI Conference on Web and Social Media, Virtual, 8–11 June 2019; Volume 14, pp. 853–862.
154. Cheng, M.; Wang, S.; Yan, X.; Yang, T.; Wang, W.; Huang, Z.; Xiao, X.; Nazarian, S.; Bogdan, P. A COVID-19 rumor dataset. *Front. Psychol.* **2021**, *12*, 644801. [CrossRef]
155. Patwa, P.; Sharma, S.; Pykl, S.; Guptha, V.; Kumari, G.; Akhtar, M.S.; Ekbal, A.; Das, A.; Chakraborty, T. Fighting an infodemic: COVID-19 fake news dataset. In Proceedings of the Combating Online Hostile Posts in Regional Languages during Emergency Situation: First International Workshop 2021, CONSTRAINT 2021, Collocated with AAAI 2021, Virtual Event, 8 February 2021; Springer International Publishing: Cham, Switzerland, 2021; pp. 21–29.
156. Yang, C.; Zhou, X.; Zafarani, R. CHECKED: Chinese COVID-19 fake news dataset. *Soc. Netw. Anal. Min.* **2021**, *11*, 58. [CrossRef] [PubMed]
157. Li, Y.; Jiang, B.; Shu, K.; Liu, H. MM-COVID: A multilingual and multimodal data repository for combating COVID-19 disinformation. *arXiv* **2020**, arXiv:2011.04088.
158. Haq, E.U.; Zia, H.B.; Mogavi, R.H.; Tyson, G.; Lu, Y.K.; Braud, T.; Hui, P. A Twitter Dataset for Pakistani Political Discourse. *arXiv* **2023**, arXiv:2301.06316.
159. Kar, D.; Bhardwaj, M.; Samanta, S.; Azad, A.P. No rumours please! a multi-indic-lingual approach for COVID fake-tweet detection. Proceedings of the 2021 Grace Hopper Celebration India (GHCI), Virtual, 19 February 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1–5.
160. Haouari, F.; Hasanain, M.; Suwaileh, R.; Elsayed, T. ArCOV19-rumors: Arabic COVID-19 twitter dataset for misinformation detection. *arXiv* **2020**, arXiv:2010.08768.
161. Alam, F.; Dalvi, F.; Shaar, S.; Durrani, N.; Mubarak, H.; Nikolov, A.; Da San Martino, G.; Abdelali, A.; Sajjad, H.; Darwish, K.; et al. Fighting the COVID-19 infodemic in social media: A holistic perspective and a call to arms. In Proceedings of the International AAAI Conference on Web and Social Media, Virtual, 7–10 June 2021; Volume 15, pp. 913–922.
162. Elhadad, M.K.; Li, K.F.; Gebali, F. COVID-19-FAKES: A Twitter (Arabic/English) dataset for detecting misleading information on COVID-19. In *Advances in Intelligent Networking and Collaborative Systems, Proceedings of the 12th International Conference on Intelligent Networking and Collaborative Systems (INCoS-2020), Victoria, BC, Canada, 31 August–2 September 2020*; Springer International Publishing: Cham, Switzerland, 2021; pp. 256–268.
163. Vlachos, A.; Riedel, S. Fact checking: Task definition and dataset construction. In Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science, Baltimore, MD, USA, 26 June 2014; pp. 18–22.
164. Yang, Y.; Zheng, L.; Zhang, J.; Cui, Q.; Li, Z.; Yu, P.S. TI-CNN: Convolutional neural networks for fake news detection. *arXiv* **2018**, arXiv:1806.00749.
165. Wang, W.Y. "Liar 2016, liar pants on fire": A new benchmark dataset for fake news detection. *arXiv* **2017**, arXiv:1705.00648.
166. Peng, X.; Xu, Q.; Feng, Z.; Zhao, H.; Tan, L.; Zhou, Y.; Zhang, Z.; Gong, C.; Zheng, Y. Automatic News Generation and Fact-Checking System Based on Language Processing. *arXiv* **2024**, arXiv:2405.10492.
167. Shrestha, A.; Spezzano, F. Textual characteristics of news title and body to detect fake news: A reproducibility study. In *Advances in Information Retrieval, Proceedings of the 43rd European Conference on IR Research 2021, ECIR 2021, Virtual Event, 28 March–1 April 2021*; Springer International Publishing: Cham, Szwitzerland, 2021; pp. 12–133.
168. Rashkin, H.; Choi, E.; Jang, J.Y.; Volkova, S.; Choi, Y. Truth of varying shades: Analyzing language in fake news and political fact-checking. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark, 9–11 September 2017; pp. 2931–2937.
169. Horne, B.; Adali, S. This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. In Proceedings of the International AAAI Conference on Web and Social Media, Montreal, QC, Canada, 15–18 May 2017; Volume 11, pp. 759–766.
170. Gandhi, R. Getting Real With Fake News, Medium. Medium. 2020. Available online: https://medium.com/riagandhi1/getting-real-with-fake-news-d4bc033eb38a (accessed on 6 January 2023).
171. Burfoot, C.; Baldwin, T. Automatic Satire Detection: Are You Having a Laugh? In Proceedings of the ACL-IJCNLP 2009 Conference Short Papers, Suntec, Singapore, 4 August 2009; pp. 161–164.
172. Pérez-Rosas, V.; Kleinberg, B.; Lefevre, A.; Mihalcea, R. Automatic detection of fake news. *arXiv* **2017**, arXiv:1708.07104.
173. Ahmed, H.; Traore, I.; Saad, S. Detection of online fake news using n-gram analysis and machine learning techniques. In *Intelligent 2017, Secure, and Dependable Systems in Distributed and Cloud Environments, Proceedings of the First International Conference, ISDDC 2017, Vancouver, BC, Canada, 26–28 October 2017*; Springer International Publishing: Cham, Switzerland, 2017; pp. 127–138.
174. Torabi Asr, F.; Taboada, M. Big Data and quality data for fake news and misinformation detection. *Big Data Soc.* **2019**, *6*, 2053951719843310. [CrossRef]
175. Zlatkova, D.; Nakov, P.; Koychev, I. Fact-checking meets fauxtography: Verifying claims about images. *arXiv* **2019**, arXiv:1908.11722.

176. Jindal, S.; Sood, R.; Singh, R.; Vatsa, M.; Chakraborty, T. Newsbag: A multimodal benchmark dataset for fake news detection. *CEUR Workshop Proc.* **2020**, *2560*, 138–145.
177. Dutta, P.S.; Das, M.; Biswas, S.; Bora, M.; Saikia, S.S. Fake news prediction: A review. *Int. J. Sci. Eng. Sci.* **2019**, *3*, 1–3.
178. Salem, F.K.A.; Al Feel, R.; Elbassuoni, S.; Jaber, M.; Farah, M. Fakes: A fake news dataset around the syrian war. In Proceedings of the International AAAI Conference on Web and Social Media, Münich, Germany, 11–14 June 2019; Volume 13, pp. 573–582.
179. Murayama, T.; Hisada, S.; Uehara, M.; Wakamiya, S.; Aramaki, E. Annotation-Scheme Reconstruction for "Fake News" and Japanese Fake News Dataset. *arXiv* **2022**, arXiv:2204.02718.
180. Posadas-Durán, J.P.; Gómez-Adorno, H.; Sidorov, G.; Escobar, J.J.M. Detection of fake news in a new corpus for the Spanish language. *J. Intell. Fuzzy Syst.* **2019**, *36*, 4869–4876. [CrossRef]
181. Amjad, M.; Sidorov, G.; Zhila, A.; Gómez-Adorno, H.; Voronkov, I.; Gelbukh, A. "Bend the truth": Benchmark dataset for fake news detection in Urdu language and its evaluation. *J. Intell. Fuzzy Syst.* **2020**, *39*, 2457–2469. [CrossRef]
182. Shahi, G.K.; Nandini, D. FakeCovid—A multilingual cross-domain fact check news dataset for COVID-19. *arXiv* **2020**, arXiv:2006.11343.
183. Thorne, J.; Vlachos, A.; Christodoulopoulos, C.; Mittal, A. Fever: A large-scale dataset for fact extraction and verification. *arXiv* **2018**, arXiv:1803.05355.
184. Thorne, J.; Vlachos, A.; Cocarascu, O.; Christodoulopoulos, C.; Mittal, A. The FEVER2. 0 shared task. In *Proceedings of the Second Workshop on Fact Extraction and VERification (FEVER)*; Association for Computational Linguistics: Hong Kong, China, 2019; pp. 1–6.
185. Aly, R.; Guo, Z.; Schlichtkrull, M.; Thorne, J.; Vlachos, A.; Christodoulopoulos, C.; Cocarascu, O.; Mittal, A. Feverous: Fact extraction and verification over unstructured and structured information. *arXiv* **2021**, arXiv:2106.05707.
186. Ferreira, W.; Vlachos, A. Emergent: A novel data-set for stance classification. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego, CA, USA, 12–17 June 2016; ACL: Washington, DC, USA, 2016.
187. Augenstein, I.; Lioma, C.; Wang, D.; Lima, L.C.; Hansen, C.; Hansen, C.; Simonsen, J.G. MultiFC: A real-world multi-domain dataset for evidence-based fact checking of claims. *arXiv* **2019**, arXiv:1909.03242.
188. Popat, K.; Mukherjee, S.; Strötgen, J.; Weikum, G. Where the truth lies: Explaining the credibility of emerging claims on the web and social media. In Proceedings of the 26th International Conference on World Wide Web Companion, Geneva,Switzerland, 3–7 April 2017; pp. 1003–1012.
189. Mishra, S.; Suryavardan, S.; Bhaskar, A.; Chopra, P.; Reganti, A.; Patwa, P.; Das, A.; Chakraborty, T.; Sheth, A.; Ekbal, A.; et al. Factify: A multi-modal fact verification dataset. In Proceedings of the First Workshop on Multimodal Fact-Checking and Hate Speech Detection (DE-FACTIFY), Virtual Event, Vancouver, BC, Canada, 27 February 2022.
190. Verma, P.K.; Agrawal, P.; Amorim, I.; Prodan, R. WELFake: Word embedding over linguistic features for fake news detection. *IEEE Trans. Comput. Soc. Syst.* **2021**, *8*, 881–893. [CrossRef]
191. Qi, P.; Bu, Y.; Cao, J.; Ji, W.; Shui, R.; Xiao, J.; Wang, D.; Chua, T.S. FakeSV: A multimodal benchmark with rich social context for fake news detection on short video platforms. *Proc. Aaai Conf. Artif. Intell.* **2023**, *37*, 14444–14452. [CrossRef]
192. Kishwar, A.; Zafar, A. Fake news detection on Pakistani news using machine learning and deep learning. *Expert Syst. Appl.* **2023**, *211*, 118558. [CrossRef]
193. Caled, D.; Carvalho, P.; Silva, M.J. MINT-Mainstream and Independent News Text Corpus. In Proceedings of the International Conference on Computational Processing of the Portuguese Language, Fortaleza, Brazil, 21–23 March 2022; Springer International Publishing: Cham, Switzerland, 2022; pp. 26–36.
194. Amjad, M.; Butt, S.; Amjad, H.I.; Zhila, A.; Sidorov, G.; Gelbukh, A. Overview of the shared task on fake news detection in Urdu at Fire. *arXiv* **2022**, arXiv:2207.05133.
195. Nørregaard, J.; Derczynski, L. DanFEVER: Claim verification dataset for Danish. In Proceedings of the 23rd Nordic Conference on Computational Linguistics, Reykjavik, Iceland (Online), 31 May–2 June 2021; pp. 422–428.
196. Min, E.; Rong, Y.; Bian, Y.; Xu, T.; Zhao, P.; Huang, J.; Ananiadou, S. Divide-and-conquer: Post-user interaction network for fake news detection on social media. In Proceedings of the ACM Web Conference 2022, Lyon, France, 25–29 April 2022; pp. 1148–1158.
197. Kumar, S.; Singh, T.D. Fake news detection on Hindi news dataset. *Glob. Transit. Proc.* **2022**, *3*, 289–297. [CrossRef]
198. Rohman, S.; Ferdous, J.; Ullah, S.M.R.; Rahman, M.A. IBFND: An Improved Dataset for Bangla Fake News Detection and Comparative Analysis of Performance of Baseline Models. In Proceedings of the 2023 International Conference on Next-Generation Computing, IoT and Machine Learning (NCIM), Gazipur, Bangladesh, 16–17 June 2023; pp. 1–6.
199. Mertoğlu, U.; Genç, B. Automated fake news detection in the age of digital libraries. *Inf. Technol. Libr.* **2020**, *39*, 4. [CrossRef]
200. Monti, F.; Frasca, F.; Eynard, D.; Mannion, D.; Bronstein, M.M. Fake news detection on social media using geometric deep learning. *arXiv* **2019**, arXiv:1902.06673v1.
201. Tompea, D. FAKE NEWS—Tool in the Information War in Ukraine. *Rev. Etica Deontol.* **2022**, *2*, 106–121. [CrossRef]
202. Oshikawa, R.; Qian, J.; Wang, W.Y. A review on natural language processing for fake news detection. *arXiv* **2020**, arXiv:1811.00770. https://arxiv.org/abs/1811.00770.
203. Pathak, A.; Srihari, R.K. BREAKING! Presenting Fake News Corpus for Automated Fact Checking. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop, Florence, Italy, 28 July–2 August 2019; Association for Computational Linguistics: Kerrville, TX, USA, 2019; pp. 357–362.

204. Shu, K.; Wang, S.; Liu, H. Beyond news contents: The role of social context for fake news detection. In Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, New York, NY, USA,30 January 2019; pp. 312–320.
205. Yuan, C.; Ma, Q.; Zhou, W.; Han, J.; Hu, S. Early Detection of Fake News by Utilizing the Credibility of News, Publishers, and Users based on Weakly Supervised Learning. In Proceedings of the 28th International Conference on Computational Linguistics, Barcelona, Spain (Online), 8–13 December 2020; pp. 5444–5454.
206. Pierri, F.; Ceri, S. False news on social media: A data-driven review. *ACM SIGMOD Rec.* **2019**, *48*, 18–27. [CrossRef]
207. Tkachenko, V. Big Dataset: All Reddit Comments–Analyzing with Clickhouse, Percona Database Performance Blog. 2017. Available online: https://www.percona.com/blog/big-data-set-reddit-comments-analyzing-clickhouse/ (accessed on 28 January 2023).
208. Pogorelov, K.; Schroeder, D.T.; Brenner, S.; Maulana, A.; Langguth, J. Combining tweets and connections graph for fake-news detection at mediaeval 2022. In Proceedings of the MediaEval'22: Multimedia Benchmark Workshop, Bergen, Norway, 13–15 January 2022.
209. Wang, Y.; Ma, F.; Jin, Z.; Yuan, Y.; Xun, G.; Jha, K.; Su, L.; Gao, J. Eann: Event adversarial neural networks for multi-modal fake news detection. In Proceedings of the 24th ACM sigkdd International Conference on Knowledge Discovery & Data Mining, London, UK, 19–23 August 2018; pp. 849–857.
210. Jin, Z.; Cao, J.; Guo, H.; Zhang, Y.; Luo, J. Multimodal fusion with recurrent neural networks for rumor detection on microblogs. In Proceedings of the 25th ACM international conference on Multimedia, New York, NY, USA, 23–27 October 2017; pp. 795–816.
211. Boididou, C.; Andreadou, K.; Papadopoulos, S.; Dang-Nguyen, D.T.; Boato, G.; Riegler, M.; Kompatsiaris, Y. Verifying multimedia use at mediaeval 2015. *MediaEval* **2015**, *3*, 7.
212. Rani, N.; Das, P.; Bhardwaj, A.K. Rumor, misinformation among web: A contemporary review of rumor detection techniques during different web waves. *Concurr. Comput. Pract. Exp.* **2022**, *34*, e6479. [CrossRef]
213. Tomaszewski, T.; Morales, A.; Lourentzou, I.; Caskey, R.; Liu, B.; Schwartz, A.; Chin, J. Identifying false human papillomavirus (HPV) vaccine information and corresponding risk perceptions from Twitter: Advanced predictive models. *J. Med. Internet Res.* **2021**, *23*, e30451. [CrossRef] [PubMed]
214. 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]
215. Willmore, A. This analysis shows how viral fake election news stories outperformed real news on Facebook. *BuzzFeed*, 16 November 2016.
216. Alhindi, T.; Petridis, S.; Muresan, S. Where is your evidence: Improving fact-checking by justification modeling. In Proceedings of the First Workshop on Fact Extraction and Verification (FEVER), Brussels, Belgium, November 2018; pp. 85–90.
217. Farmer, S.J.L. False, Misleading, Clickbait-Y, and Satirical "News" sources, Fake News in Context. 2016. Available online: https://d279m997dpfwgl.cloudfront.net/wp/2016/11/Resource-False-Misleading-Clickbait-y-and-Satirica80Sources-1.pdf (accessed on 6 January 2023).
218. Hanselowski, A.; Stab, C.; Schulz, C.; Li, Z.; Gurevych, I. A richly annotated corpus for different tasks in automated fact-checking. *arXiv* **2019**, arXiv:1911.01214.
219. Godel, W.; Sanderson, Z.; Aslett, K.; Nagler, J.; Bonneau, R.; Persily, N.; Tucker, J.A. Moderating with the mob: Evaluating the efficacy of real-time crowdsourced fact-checking. *J. Online Trust. Saf.* **2021**, *1*. [CrossRef]
220. Espina Mairal, S.; Bustos, F.; Solovey, G.; Navajas, J. Interactive crowdsourcing to fact-check politicians. *J. Ex Psychol. Appl.* **2024**, *30*, 3. [CrossRef]
221. Shim, J.S.; Lee, Y.; Ahn, H. A link2vec-based fake news detection model using web search results. *Expert Syst. Appl.* **2021**, *184*, 115491. [CrossRef]
222. Sheikhi, S. An effective fake news detection method using WOA-xgbTree algorithm and content-based features. *Appl. Soft Comput.* **2021**, *109*, 107559. [CrossRef]
223. Ngada, O.; Haskins, B. Fake news detection using content-based features and machine learning. In Proceedings of the 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), Gold Coast, Australia, 16–18 December 2020; pp. 1–6.
224. Kshetri, N.; Voas, J. The economics of "fake news". *IT Prof.* **2017**, *19*, 8–12. [CrossRef]
225. Zhang, X.; Lashkari, A.H.; Ghorbani, A.A. A Lightweight Online Advertising Classification System using Lexical-based Features. In Proceedings of the 14th International Joint Conference on e-Business and Telecommunications (ICETE 2017) SECRYPT, Madrid, Spain, 24–26 July 2017; pp. 486–494.
226. Alrubaian, M.; Al-Qurishi, M.; Hassan, M.M.; Alamri, A. A credibility analysis system for assessing information on twitter. *IEEE Trans. Dependable Secur. Comput.* **2016**, *15*, 661–674. [CrossRef]
227. Arin, E.; Kutlu, M. Deep learning based social bot detection on twitter. *IEEE Trans. Inf. Forensics Secur.* **2023**, *18*, 1763–1772. [CrossRef]
228. Sahoo, S.R.; Gupta, B.B.; Peraković, D.; Peñalvo, F.J.G.; Cvitić, I. Spammer detection approaches in online social network (OSNs): A review. In *Sustainable Management of Manufacturing Systems in Industry 4.0*; Springer International Publishing: Cham, Switzerland, 2022; pp. 159–180.
229. Khaund, T.; Kirdemir, B.; Agarwal, N.; Liu, H.; Morstatter, F. Social bots and their coordination during online campaigns: A review. *IEEE Trans. Comput. Soc. Syst.* **2021**, *9*, 530–545. [CrossRef]

230. Shahbazi, Z.; Byun, Y.C. Fake media detection based on natural language processing and blockchain approaches. *IEEE Access* **2021**, *9*, 128442–128453. [CrossRef]
231. Pang, B.; Lee, L. Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.* **2008**, *2*, 1–135. [CrossRef]
232. Costola, M.; Hinz, O.; Nofer, M.; Pelizzon, L. Machine learning sentiment analysis, COVID-19 news and stock market reactions. *Res. Int. Bus. Financ.* **2023**, *64*, 101881. [CrossRef]
233. Dodds, P.S.; Clark, E.M.; Desu, S.; Frank, M.R.; Reagan, A.J.; Williams, J.R.; Mitchell, L.; Harris, K.D.; Kloumann, I.M.; Bagrow, J.P.; et al. Human language reveals a universal positivity bias. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 2389–2394. [CrossRef]
234. Cheng, Y.; Yao, L.; Xiang, G.; Zhang, G.; Tang, T.; Zhong, L. Text sentiment orientation analysis based on multi-channel CNN and bidirectional GRU with attention mechanism. *IEEE Access* **2020**, *8*, 134964–134975. [CrossRef]
235. Chen, Y.; Li, D.; Zhang, P.; Sui, J.; Lv, Q.; Tun, L.; Shang, L. Cross-modal ambiguity learning for multimodal fake news detection. In Proceedings of the ACM Web Conference, New York, NY, USA, 25–29 April 2022; pp. 2897–2905.
236. Ying, Q.; Hu, X.; Zhou, Y.; Qian, Z.; Zeng, D.; Ge, S. Bootstrapping Multi-view Representations for Fake News Detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*; AAAI: Washington, DC, USA, 2023.
237. Suryavardan, S.; Mishra, S.; Chakraborty, M.; Patwa, P.; Rani, A.; Chadha, A.; Reganti, A.; Das, A.; Sheth, A.; Chinnakotla, M.; et al. Findings of factify 2: Multimodal fake news detection. *arXiv* **2023**, arXiv:2307.10475.
238. Zhou, Y.; Yang, Y.; Ying, Q.; Qian, Z.; Zhang, X. Multimodal fake news detection via clip-guided learning. In Proceedings of the 2023 IEEE International Conference on Multimedia and Expo (ICME), Brisbane, Australia, 10–14 July 2023; pp. 2825–2830.
239. Jin, Z.; Cao, J.; Zhang, Y.; Zhou, J.; Tian, Q. Novel visual and statistical image features for microblogs news verification. *IEEE Trans. Multimed.* **2016**, *19*, 598–608. [CrossRef]
240. Giachanou, A.; Zhang, G.; Rosso, P. Multimodal multi-image fake news detection. In Proceedings of the 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA), Sydney, Australia, 6–9 October 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 647–654.
241. Singhal, S.; Kabra, A.; Sharma, M.; Shah, R.R.; Chakraborty, T.; Kumaraguru, P. Spotfake+: A multimodal framework for fake news detection via transfer learning (student abstract). In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 13915–13916.
242. Shah, P.; Kobti, Z. Multimodal fake news detection using a Cultural Algorithm with situational and normative knowledge. In Proceedings of the 2020 IEEE Congress on Evolutionary Computation (CEC), Glasgow, UK, 19–24 July 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–7.
243. Segura-Bedmar, I.; Alonso-Bartolome, S. Multimodal fake news detection. *Information* **2022**, *13*, 284. [CrossRef]
244. Singhal, S.; Shah, R.R.; Chakraborty, T.; Kumaraguru, P.; Satoh, S.I. Spotfake: A multi-modal framework for fake news detection. In Proceedings of the 2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM), Singapore, 11–13 September 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 39–47.
245. Tanwar, V.; Sharma, K. Multi-model fake news detection based on concatenation of visual latent features. In Proceedings of the 2020 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 28–30 July 2020; IEEE: Piscataway, NJ, USA, 2020; p. 13441348.
246. Singh, V.K.; Ghosh, I.; Sonagara, D. Detecting fake news stories via multimodal analysis. *J. Assoc. Inf. Sci. Technol.* **2021**, *72*, 3–17. [CrossRef]
247. Choudhury, D.; Acharjee, T. A novel approach to fake news detection in social networks using genetic algorithm applying machine learning classifiers. *Multimed. Tools Appl.* **2023**, *82*, 9029–9045. [CrossRef]
248. Xu, J.; Li, Z.; Huang, F.; Li, C.; Philip, S.Y. Visual sentiment analysis with social relations-guided multiattention networks. *IEEE Trans. Cybern.* **2020**, *52*, 4472–4484. [CrossRef] [PubMed]
249. Sansonetti, G.; Gasparetti, F.; D'aniello, G.; Micarelli, A. Unreliable users detection in social media: Deep learning techniques for automatic detection. *IEEE Access* **2020**, *8*, 213154–213167. [CrossRef]
250. Kaur, S.; Kumar, P.; Kumaraguru, P. Automating fake news detection system using multi-level voting model. *Soft Comput.* **2020**, *24*, 9049–9069. [CrossRef]
251. Jiang, T.A.O.; Li, J.P.; Haq, A.U.; Saboor, A.; Ali, A. A novel stacking approach for accurate detection of fake news. *IEEE Access* **2021**, *9*, 22626–22639. [CrossRef]
252. Granik, M.; Mesyura, V. Fake news detection using naive Bayes classifier. In Proceedings of the 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), Kyiv, Ukraine, 29 May–2 June 2017; pp. 900–903.
253. Kotteti, C.M.M.; Dong, X.; Li, N.; Qian, L. Fake news detection enhancement with data imputation. In Proceedings of the 2018 IEEE 16th International Conference on Dependable, Autonomic and Secure Computing, 16th International Conference on Pervasive Intelligence and Computing, 4th International Conference on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech), Athens, Greece, 12–15 August 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 187–192.
254. Ni, S.; Li, J.; Kao, H.Y. MVAN: Multi-view attention networks for fake news detection on social media. *IEEE Access* **2021**, *9*, 106907–106917. [CrossRef]
255. Fayaz, M.; Khan, A.; Bilal, M.; Khan, S.U. Machine learning for fake news classification with optimal feature selection. *Soft Comput.* **2022**, *26*, 7763–7771. [CrossRef]

256. Kesarwani, A.; Chauhan, S.S.; Nair, A.R. Fake news detection on social media using k-nearest neighbor classifier. In Proceedings of the 2020 International Conference on Advances in Computing and Communication Engineering (ICACCE), Las Vegas, NV, USA, 22–24 June 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–4.
257. Wynne, H.E.; Wint, Z.Z. Content based fake news detection using n-gram models. In Proceedings of the 21st International Conference on Information Integration and Web-Based Applications & Services, New York, NY, USA, 2–4 December 2019; pp. 669–673.
258. Ganesh, P.; Priya, L.; Nandakumar, R. Fake news detection-a comparative study of advanced ensemble approaches. In Proceedings of the 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 3–5 June 2021; pp. 1003–1008.
259. Jain, M.K.; Gopalani, D.; Meena, Y.K.; Kumar, R. Machine Learning based Fake News Detection using linguistic features and word vector features. In Proceedings of the 2020 IEEE 7th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), Prayagraj, India, 27–29 November 2020; pp. 1–6.
260. Nasir, J.A.; Khan, O.S.; Varlamis, I. Fake news detection: A hybrid CNN-RNN based deep learning approach. *Int. J. Inf. Manag. Data Insights* **2021**, *1*, 100007. [CrossRef]
261. Tai, K.Y.; Dhaliwal, J.; Shariff, S.M. Online social networks and writing styles—A review of the multidisciplinary literature. *IEEE Access* **2020**, *8*, 67024–67046. [CrossRef]
262. Wang, Y.; Li, B. Sentiment analysis for social media images. In Proceedings of the 2015 IEEE International Conference on Data Mining Workshop (ICDMW), Atlantic City, NJ, USA, 14–17 November 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 1584–1591.
263. Keramatfar, A.; Amirkhani, H.; Bidgoly, A.J. Modeling tweet dependencies with graph convolutional networks for sentiment analysis. *Cogn. Comput.* **2022**, *14*, 2234–2245. [CrossRef]
264. Minh, H.L.; Sang-To, T.; Wahab, M.A.; Cuong-Le, T. A new metaheuristic optimization based on K-means clustering algorithm and its application to structural damage identification. *Knowl.-Based Syst.* **2022**, *251*, 109189. [CrossRef]
265. Lu, X.S.; Zhou, M.; Qi, L.; Liu, H. Clustering-algorithm-based rare-event evolution analysis via social media data. *IEEE Trans. Comput. Soc. Syst.* **2019**, *6*, 301–310. [CrossRef]
266. Naredla, N.R.; Adedoyin, F.F. Detection of hyperpartisan news articles using natural language processing technique. *Int. J. Inf. Manag. Data Insights* **2022**, *2*, 100064. [CrossRef]
267. Bojanowski, P.; Grave, E.; Joulin, A.; Mikolov, T. Enriching word vectors with subword information. *Trans. Assoc. Comput. Linguist.* **2017**, *5*, 135–146. [CrossRef]
268. Li, D.; Guo, H.; Wang, Z.; Zheng, Z. Unsupervised fake news detection based on autoencoder. *IEEE Access* **2021**, *9*, 29356–29365. [CrossRef]
269. de Souza, M.C.; Nogueira, B.M.; Rossi, R.G.; Marcacini, R.M.; Dos Santos, B.N.; Rezende, S.O. A network-based positive and unlabeled learning approach for fake news detection. *Mach. Learn.* **2022**, *111*, 3549–3592. [CrossRef]
270. Huang, Y.F.; Chen, P.H. Fake news detection using an ensemble learning model based on self-adaptive harmony search algorithms. *Expert Syst. Appl.* **2020**, *159*, 113584. [CrossRef]
271. Sánchez-Fernández, P.; Ruiz, L.G.B.; Jiménez, M.D.C.P. Application of classical and advanced machine learning models to predict personality on social media. *Expert Syst. Appl.* **2023**, *216*, 119498. [CrossRef]
272. Reddy, H.; Raj, N.; Gala, M.; Basava, A. Text-mining-based fake news detection using ensemble methods. *Int. J. Autom. Comput.* **2020**, *17*, 210–221. [CrossRef]
273. 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]
274. Mahabub, A. A robust technique of fake news detection using Ensemble Voting Classifier and comparison with other classifiers. *SN Appl. Sci.* **2020**, *2*, 525. [CrossRef]
275. Deng, R.; Duzhin, F. Topological data analysis helps to improve accuracy of deep learning models for fake news detection trained on very small training sets. *Big Data Cogn. Comput.* **2022**, *6*, 74. [CrossRef]
276. Lee, D.H.; Kim, Y.R.; Kim, H.J.; Park, S.M.; Yang, Y.J. Fake news detection using deep learning. *J. Inf. Process. Syst.* **2019**, *15*, 1119–1130.
277. Dong, X.; Victor, U.; Qian, L. Two-path deep semisupervised learning for timely fake news detection. *IEEE Trans. Comput. Soc. Syst.* **2020**, *7*, 1386–1398. [CrossRef]
278. Van Houdt, G.; Mosquera, C.; Nápoles, G. A review on the long short-term memory model. *Artif. Intell. Rev.* **2020**, *53*, 5929–5955. [CrossRef]
279. Dong, X.; Qian, L. Semi-supervised bidirectional RNN for misinformation detection. *Mach. Learn. Appl.* **2022**, *10*, 100428. [CrossRef]
280. Bahad, P.; Saxena, P.; Kamal, R. Fake news detection using bi-directional LSTM-recurrent neural network. *Procedia Comput. Sci.* **2019**, *165*, 74–82. [CrossRef]
281. Long, Y.; Lu, Q.; Xiang, R.; Li, M.; Huang, C.R. Fake news detection through multi-perspective speaker profiles. In Proceedings of the Eighth International Joint Conference on Natural Language Processing, Taipei, Taiwan, 27 November– 1 December 2017; Volume 2.
282. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.

283. Umer, M.; Imtiaz, Z.; Ullah, S.; Mehmood, A.; Choi, G.S.; On, B.W. Fake news stance detection using deep learning architecture (CNN-LSTM). *IEEE Access* **2020**, *8*, 156695–156706. [CrossRef]
284. Upadhyay, R.; Pasi, G.; Viviani, M. Vec4Cred: A model for health misinformation detection in web pages. *Multimed. Tools Appl.* **2023**, *82*, 5271–5290. [CrossRef]
285. Li, P.; Sun, X.; Yu, H.; Tian, Y.; Yao, F.; Xu, G. Entity-oriented multi-modal alignment and fusion network for fake news detection. *IEEE Trans. Multimed.* **2021**, *24*, 3455–3468. [CrossRef]
286. Galende, B.A.; Hernández-Peñaloza, G.; Uribe, S.; García, F.Á. Conspiracy or not? A deep learning approach to spot it on Twitter. *IEEE Access* **2022**, *10*, 38370–38378. [CrossRef]
287. Kurniasari, L.; Setyanto, A. Sentiment analysis using recurrent neural network. *J. Phys. Conf. Ser.* **2020**, *1471*, 012018. [CrossRef]
288. Yang, M.; Liu, S.; Chen, K.; Zhang, H.; Zhao, E.; Zhao, T. A hierarchical clustering approach to fuzzy semantic representation of rare words in neural machine translation. *IEEE Trans. Fuzzy Syst.* **2020**, *28*, 992–1002. [CrossRef]
289. Le, T.; Wang, S.; Lee, D. Malcom: Generating malicious comments to attack neural fake news detection models. In Proceedings of the 2020 IEEE International Conference on Data Mining (ICDM), Sorrento, Italy, 17–20 November 2020; pp. 282–291.
290. Kaliyar, R.K. Fake news detection using a deep neural network. In Proceedings of the 2018 4th International Conference on Computing Communication and Automation (ICCCA), Greater Noida, India, 14–15 December 2018; pp. 1–7.
291. Nadeem, M.I.; Ahmed, K.; Li, D.; Zheng, Z.; Alkahtani, H.K.; Mostafa, S.M.; Mamyrbayev, O.; Abdel Hameed, H. EFND: A semantic, visual, and socially augmented deep framework for extreme fake news detection. *Sustainability* **2022**, *15*, 133. [CrossRef]
292. Gautam, R.; Sharma, M. Prevalence and diagnosis of neurological disorders using different deep learning techniques: A meta-analysis. *J. Med. Syst.* **2020**, *44*, 49. [CrossRef]
293. Muthu Lakshmi, V.; Radhika, R.; Kalpana, G. Radial Restricted Boltzmann Machines with Functional Neural Network for Classification of the Fake and Real News Analysis. *Int. J. Uncertain. Fuzziness-Knowl.-Based Syst.* **2022**, *30* (Suppl. S1), 31–43. [CrossRef]
294. Cao, W.; Yan, Z.; He, Z.; He, Z. A comprehensive review on geometric deep learning. *IEEE Access* **2020**, *8*, 35929–35949. [CrossRef]
295. Wang, Z.; Dong, Q.; Guo, W.; Li, D.; Zhang, J.; Du, W. Geometric imbalanced deep learning with feature scaling and boundary sample mining. *Pattern Recognit.* **2022**, *126*, 108564. [CrossRef]
296. Villalba-Diez, J.; Molina, M.; Schmidt, D. Geometric Deep Lean Learning: Evaluation Using a Twitter Social Network. *Appl. Sci.* **2021**, *11*, 6777. [CrossRef]
297. Collobert, R.; Weston, J.; Bottou, L.; Karlen, M.; Kavukcuoglu, K.; Kuksa, P. Natural language processing (almost) from scratch. *J. Mach. Learn. Res.* **2011**, *12*, 2493–2537.
298. Sarkar, S.; Tudu, N.; Das, D. HIJLI-JU-CLEF at MULTI-Fake-DetectiVE: Multimodal Fake News Detection Using Deep Learning Approach. In Proceedings of the Eighth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2023), Parma, Italy, 7–8 September 2023.
299. Verma, P.K.; Agrawal, P.; Madaan, V.; Prodan, R. MCred: Multi-modal message credibility for fake news detection using BERT and CNN. *J. Ambient Intell. Humaniz. Comput.* **2023**, *14*, 10617–10629. [CrossRef] [PubMed]
300. Praseed, A.; Rodrigues, J.; Thilagam, P.S. Hindi fake news detection using transformer ensembles. *Eng. Appl. Artif. Intell.* **2023**, *119*, 105731. [CrossRef]
301. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
302. Wang, B.; Feng, Y.; Xiong, X.C.; Wang, Y.H.; Qiang, B.H. Multi-modal transformer using two-level visual features for fake news detection. *Appl. Intell.* **2023**, *53*, 10429–10443. [CrossRef]
303. Kenton, J.D.M.W.C.; Toutanova, L.K. Bert: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the NAACL-HLT, Minneapolis, MN, USA, 2–7 June 2019; Volumes 1–2.
304. Wang, A.; Singh, A.; Michael, J.; Hill, F.; Levy, O.; Bowman, S.R. GLUE: A multi-task benchmark and analysis platform for natural language understanding. *arXiv* **2018**, arXiv:1804.07461.
305. Baarir, N.F.; Djeffal, A. Fake news detection using machine learning. In Proceedings of the 2020 2nd International Workshop on Human-Centric Smart Environments for Health and Well-Being (IHSH), Boumerdes, Algeria, 9–10 February 2021; pp. 125–130.
306. Shrivastava, S.; Singh, R.; Jain, C.; Kaushal, S. A research on fake news detection using machine learning algorithm. In *Smart Systems: Innovations in Computing, Proceedings of the SSIC 2021*; Springer: Singapore, 2022; pp. 273–287.
307. Agarwal, A.; Mittal, M.; Pathak, A.; Goyal, L.M. Fake news detection using a blend of neural networks: An application of deep learning. *SN Comput. Sci.* **2020**, *1*, 143. [CrossRef]
308. Saleh, H.; Alharbi, A.; Alsamhi, S.H. OPCNN-FAKE: Optimized convolutional neural network for fake news detection. *IEEE Access* **2021**, *9*, 129471–129489. [CrossRef]
309. Kaliyar, R.K.; Goswami, A.; Narang, P. DeepFakE: Improving fake news detection using tensor decomposition-based deep neural network. *J. Supercomput.* **2021**, *77*, 1015–1037. [CrossRef]
310. 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]
311. Choudhary, M.; Chouhan, S.S.; Pilli, E.S.; Vipparthi, S.K. BerConvoNet: A deep learning framework for fake news classification. *Appl. Soft Comput.* **2021**, *110*, 107614. [CrossRef]

312. Trueman, T.E.; Kumar, A.; Narayanasamy, P.; Vidya, J. Attention-based C-BiLSTM for fake news detection. Appl. *Soft Comput.* **2021**, *110*, 107600. [CrossRef]
313. Kumari, R.; Ekbal, A. Amfb: Attention based multimodal factorized bilinear pooling for multimodal fake news detection. *Expert Syst. Appl.* **2021**, *184*, 115412. [CrossRef]
314. Lample, G.; Conneau, A. Cross-lingual language model pretraining. *arXiv* **2019**, arXiv:1901.07291.
315. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 1–11.
316. Schütz, M.; Schindler, A.; Siegel, M.; Nazemi, K. Automatic fake news detection with pre-trained transformer models. In Pattern Recognition. ICPR International Workshops and Challenges: Virtual Event, 10–15 January 2021; Springer International Publishing: Cham, Switzerland, 2021; pp. 627–641.
317. Guo, Z.; Zhang, Q.; Ding, F.; Zhu, X.; Yu, K. A Novel Fake News Detection Model for Context of Mixed Languages Through Multiscale Transformer. *IEEE Trans. Comput. Soc. Syst.* **2023**, *99*, 5079–5089. [CrossRef]
318. Raffel, C.; Shazeer, N.; Roberts, A.; Lee, K.; Narang, S.; Matena, M.; Zhou, Y.; Li, W.; Liu, P.J. Exploring the limits of transfer learning with a unified textto-text transformer. *arXiv* **2019**, arXiv1910.10683.
319. Le, H.; Vial, L.; Frej, J.; Segonne, V.; Coavoux, M.; Lecouteux, B.; Allauzen, A.; Crabbé, B.; Besacier, L.; Schwab, D. Flaubert: Unsupervised language model pre-training for french. *arXiv* **2019**, arXiv:1912.05372.
320. Subakan, C.; Ravanelli, M.; Cornell, S.; Bronzi, M.; Zhong, J. Attention is all you need in speech separation. In Proceedings of the ICASSP 2021—2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 6–11 June 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 21–25.
321. Li, J.; Hui, B.; Cheng, R.; Qin, B.; Ma, C.; Huo, N.; Huang, F.; Du, W.; Si, L.; Li, Y. Graphix-t5: Mixing pre-trained transformers with graph-aware layers for text-to-sql parsing. *arXiv* **2023**, arXiv:2301.07507. [CrossRef]
322. Yang, Z.; Dai, Z.; Yang, Y.; Carbonell, J.; Salakhutdinov, R.R.; Le, Q.V. Xlnet: Generalized autoregressive pretraining for language understanding. *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 5753–5763.
323. Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; Stoyanov, V. Roberta: A robustly optimized bert pretraining approach. *arXiv* **2019**, arXiv:1907.11692.
324. Gokaslan, A.; Cohen, V.; Pavlick, E.; Tellex, S. Openwebtext Corpus. 2019. Available online: https://skylion007.github.io/OpenWebTextCorpus/ (accessed on 20 April 2024).
325. Trinh, T.H.; Le, Q.V. A simple method for commonsense reasoning. *arXiv* **2018**, arXiv:1806.02847.
326. Rajpurkar, P.; Zhang, J.; Lopyrev, K.; Liang, P. Squad: 100,000+ questions for machine comprehension of text. *arXiv* **2016**, arXiv:1606.05250.
327. Sanh, V.; Debut, L.; Chaumond, J.; Wolf, T. DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter. *arXiv* **2019**, arXiv:1910.01108.
328. Hinton, G.; Vinyals, O.; Dean, J. Distilling the knowledge in a neural network. *arXiv* **2015**, arXiv:1503.02531.
329. Lan, Z.; Chen, M.; Goodman, S.; Gimpel, K.; Sharma, P.; Soricut, R. Albert: A lite bert for self-supervised learning of language representations. *arXiv* **2019**, arXiv:1909.11942.
330. Conneaut, A.; Khandelwal, K.; Goyal, N.; Chaudhary, V.; Wenzek, G.; Guzman, F.; Grave, E.; Ott, M.; Zettlemoyer, L.; Stoyanov, V. Unsupervised Cross-lingual Representation Learning at Scale. In Proceedings of the ACL 2020, Online, 5–10 July 2020.
331. Schütz, M.; Böck, J.; Andresel, M.; Kirchknopf, A.; Liakhovets, D.; Slijepčević, D.; Schindler, A. AIT FHSTP at CheckThat! 2022: Cross-lingual fake news detection with a large pre-trained transformer. Working Notes of CLEF. In Proceedings of the CLEF 2022: Conference and Labs of the Evaluation Forum, Bologna, Italy, 5–8 September 2022.
332. Lewis, M.; Liu, Y.; Goyal, N.; Ghazvininejad, M.; Mohamed, A.; Levy, O.; Stoyanov, V.; Zettlemoyer, L. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv* **2019**, arXiv:1910.13461.
333. Liu, Y.; Gu, J.; Goyal, N.; Li, X.; Edunov, S.; Ghazvininejad, M.; Lewis, M.; Zettlemoyer, L. Multilingual denoising pre-training for neural machine translation. *Trans. Assoc. Comput. Linguist.* **2020**, *8*, 726–742. [CrossRef]
334. Clark, K.; Luong, M.T.; Le, Q.V.; Manning, C.D. Electra: Pre-training text encoders as discriminators rather than generators. *arXiv* **2020**, arXiv:2003.10555.
335. Radford, A.; Narasimhan, K.; Salimans, T.; Sutskever, I. Improving Language Understanding by Generative Pre-Training. 2018. Available online: https://www.semanticscholar.org/paper/Improving-Language-Understanding-by-Generative-Radford-Narasimhan/cd18800a0fe0b668a1cc19f2ec95b5003d0a5035 (accessed on 20 April 2024).
336. Radford, A.; Wu, J.; Child, R.; Luan, D.; Amodei, D.; Sutskever, I. Language models are unsupervised multitask learners. *OpenAI Blog* **2019**, *1*, 9.
337. Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language models are few-shot learners. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 1877–1901.
338. Caramancion, K.M. News Verifiers Showdown: A Comparative Performance Evaluation of ChatGPT 3.5, ChatGPT 4.0, Bing AI, and Bard in News Fact-Checking. *arXiv* **2023**, arXiv:2306.17176.
339. Guo, B.; Zhang, X.; Wang, Z.; Jiang, M.; Nie, J.; Ding, Y.; Yue, J.; Wu, Y. How close is chatgpt to human experts? Comparison corpus, evaluation, and detection. *arXiv* **2023**, arXiv:2301.07597.
340. Beltagy, I.; Peters, M.E.; Cohan, A. Longformer: The long-document transformer. *arXiv* **2020**, arXiv:2004.05150.
341. Kitaev, N.; Kaiser, Ł.; Levskaya, A. Reformer: The efficient transformer. *arXiv* **2020**, arXiv:2001.04451.

342. Zaheer, M.; Guruganesh, G.; Dubey, K.A.; Ainslie, J.; Alberti, C.; Ontanon, S.; Pham, P.; Ravula, A.; Wang, Q.; Yang, L.; et al. Big bird: Transformers for longer sequences. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 17283–17297.
343. Dai, Z.; Lai, G.; Yang, Y.; Le, Q. Funnel-transformer: Filtering out sequential redundancy for efficient language processing. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 4271–4282.
344. Zhang, J.; Zhao, Y.; Saleh, M.; Liu, P. Pegasus: Pre-training with extracted gap-sentences for abstractive summarization. In Proceedings of the International Conference on Machine Learning, PMLR, Online, 21 November 2020; pp. 11328–11339.
345. Keskar, N.S.; McCann, B.; Varshney, L.R.; Xiong, C.; Socher, R. Ctrl: A conditional transformer language model for controllable generation. *arXiv* **2019**, arXiv:1909.05858.
346. Cui, Y.; Che, W.; Liu, T.; Qin, B.; Yang, Z. Pre-training with whole word masking for Chinese bert. In *Proceedings of the IEEE/ACM Transactions on Audio 2021, Speech, and Language Processing*; IEEE Press: Piscataway, NJ, USA, 2021; Volume 29, pp. 3504–3514.
347. Fathullah, Y.; Wu, C.; Lakomkin, E.; Jia, J.; Shangguan, Y.; Li, K.; Guo, J.; Xiong, W.; Mahadeokar, J.; Kalinli, O.; et al. Prompting large language models with speech recognition abilities. *arXiv* **2023**, arXiv:2307.11795.
348. Touvron, H.; Lavril, T.; Izacard, G.; Martinet, X.; Lachaux, M.A.; Lacroix, T.; Rozière, B.; Goyal, N.; Hambro, E.; Azhar, F.; et al. Llama: Open and efficient foundation language models. *arXiv* **2023**, arXiv:2302.13971.
349. Floridi, L. AI as agency without intelligence: On ChatGPT, large language models, and other generative models. *Philos. Technol.* **2023**, *36*, 15. [CrossRef]
350. Yang, R.; Tan, T.F.; Lu, W.; Thirunavukarasu, A.J.; Ting, D.S.W.; Liu, N. Large language models in health care: Development, applications, and challenges. *Health Care Sci.* **2023**, *2*, 255–263. [CrossRef]
351. Fan, L.; Li, L.; Ma, Z.; Lee, S.; Yu, H.; Hemphill, L. A bibliometric review of large language models research from 2017 to 2023. *arXiv* **2023**, arXiv:2304.02020. [CrossRef]
352. Jayaseelan, N. LLaMA 2: The New Open Source Language Model. Available online: https://www.e2enetworks.com/blog/llama-2-the-new-open-source-language-model (accessed on 20 April 2024).
353. Huang, J.; Chang, K.C.C. Citation: A key to building responsible and accountable large language models. *arXiv* **2023**, arXiv:2307.02185.
354. Pan, Y.; Pan, L.; Chen, W.; Nakov, P.; Kan, M.Y.; Wang, W.Y. On the Risk of Misinformation Pollution with Large Language Models. *arXiv* **2023**, arXiv:2305.13661.
355. De Angelis, L.; Baglivo, F.; Arzilli, G.; Privitera, G.P.; Ferragina, P.; Tozzi, A.E.; Rizzo, C. ChatGPT and the rise of large language models: The new AI-driven infodemic threat in public health. *Front. Public Health* **2023**, *11*, 1166120. [CrossRef] [PubMed]
356. Thoppilan, R.; De Freitas, D.; Hall, J.; Shazeer, N.; Kulshreshtha, A.; Cheng, H.T.; Jin, A.; Bos, T.; Baker, L.; Du, Y.; et al. Lamda: Language models for dialog applications. *arXiv* **2022**, arXiv:2201.08239.

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