future internet

MDPI

Article

Cross-Domain Fake News Detection Through Fusion of Evidence from Multiple Social Media Platforms

Jannatul Ferdush [1], Joarder Kamruzzaman [1,*], Gour Karmakar [1], Iqbal Gondal [2] and Rajkumar Das [1]

[1] Centre for Smart Analytics, Institute of Innovation, Science and Sustainability, Federation University, Ballarat, VIC 3353, Australia; jannatulferdush@students.federation.edu.au (J.F.); gour.karmakar@federation.edu.au (G.K.); r.das@federation.edu.au (R.D.)

[2] School of Computing Technologies, STEM College, RMIT University, Melbourne, VIC 3000, Australia; iqbal.gondal@rmit.edu.au

* Correspondence: joarder.kamruzzaman@federation.edu.au

Abstract: Fake news has become a significant challenge on online social platforms, increasing uncertainty and unwanted tension in society. The negative impact of fake news on political processes, public health, and social harmony underscores the urgency of developing more effective detection systems. Existing methods for fake news detection often focus solely on one platform, potentially missing important clues that arise from multiple platforms. Another important consideration is that the domain of fake news changes rapidly, making cross-domain analysis more difficult than in-domain analysis. To address both of these limitations, our method takes evidence from multiple social media platforms, enhances our cross-domain analysis, and improves overall detection accuracy. Our method employs the Dempster–Shafer combination rule for aggregating probabilities for comments being fake from two different social media platforms. Instead of directly using the comments as features, our approach improves fake news detection by examining the relationships and calculating correlations among comments from different platforms. This provides a more comprehensive view of how fake news spreads and how users respond to it. Most importantly, our study reveals that true news is typically rich in content, while fake news tends to generate a vast thread of comments. Therefore, we propose a combined method that merges content- and comment-based approaches, allowing our model to identify fake news with greater accuracy and showing an overall improvement of 7% over previous methods.

Keywords: fake news; comment; uncertainty; Dempster–Shafer theory; social media

check for updates

Academic Editors: Jinpeng Chen, Ruifan Li and Kaimin Wei

Received: 7 December 2024
Revised: 17 January 2025
Accepted: 20 January 2025
Published: 3 February 2025

Citation: Ferdush, J.; Kamruzzaman, J.; Karmakar, G.; Gondal, I.; Das, R. Cross-Domain Fake News Detection Through Fusion of Evidence from Multiple Social Media Platforms. *Future Internet* **2025**, *17*, 61. https://doi.org/10.3390/fi17020061

1. Introduction

In today's digital age, social media has transformed how we access and consume news. News is now accessible to anyone with internet access, can be shared almost instantly, and reaches millions of users worldwide within seconds. However, along with the benefits of real-time information sharing comes the rise of fake news, which can have far-reaching negative consequences for individuals, communities, and even entire nations [1–5]. The ubiquity of fake news poses significant risks to democratic practices by distorting the flow of accurate information necessary for informed citizenry, creating false narratives, and undermining legitimate journalism [6,7]. It has also demonstrated the capacity to incite public panic, influence election outcomes, and destabilize financial markets [8,9]. The importance of combating fake news is thus not just a matter of information integrity but also a requisite for safeguarding democracy, public safety, and economic stability. However, social media makes it more difficult to detect fake news. A study [10] led by the University of Southern California (USC) of over 2400 Facebook users found that social media platforms, not individual users, have a big responsibility in stopping fake news. They discovered that the way these platforms are designed encourages people to continue sharing information as they receive it without checking, which leads to further spread of fake news. Another important factor is the absence of an editorial body—which is an important component in mainstream media.

In the area of fake news detection by machine learning, traditional methodologies often limit their analysis to one domain, under the assumption that future news will always align with the domain of their training data. This is known as in-domain analysis. However, this assumption often fails to identify future fake news accurately due to the fluid nature of news topics and the differing language across domains [11]. For instance, a machine learning model trained on political news may encounter challenges when faced with fake news related to health crises such as COVID-19. Its efficacy is compromised due to the incongruity in vocabulary and news topics conventions between political and health-related news. Whereas political news typically encompasses discussions surrounding geopolitical tensions and societal affairs, COVID-19-related content emphasizes terms such as vaccines, death rates, and viral spread. Therefore, the significance of cross-domain analysis becomes apparent, highlighting its indispensable role in enhancing the efficacy of fake news detection methodologies. While some content-based articles do attempt cross-domain analysis, their detection accuracy is not so high. On the other hand, focusing on single social media platforms is always biased because different social media platforms attract different groups of people for their unique characteristics. For example, while younger users may frequent platforms like TikTok or Instagram [12], older people are more like to use Facebook [13]. People who are really into politics use Twitter more because it is good for sharing news/opinions quickly [14,15]. LinkedIn might be useful for professionals looking for work-related information [16]. Additionally, not only do different user groups favor different social media platforms, but various geographical regions also have their own preferred platforms. For example, Facebook and Twitter are commonly used in the United States alongside platforms like Reddit and YouTube, which attract a global yet often English-speaking audience [17,18]. In contrast, VKontakte (VK) is more popular in Russia [19]. In China, where many Western platforms are restricted, local services like Weibo and WeChat are most popular [20]. LINE has a significant user base in Japan [21]. Therefore, relying solely on data from a single social media platform can lead to biased or incomplete conclusions for detecting fake news.

The fluid nature of news and the rapid generation of data on social media, coupled with the current low detection accuracy in the literature [22–26], make cross-domain research an important topic. Although the literature contains some studies on cross-domain fake news detection, most of these works focus predominantly on news text, overlooking an important element of social media—the user comments. While the concept of detecting fake news through user comments is not entirely novel, existing studies in this domain have primarily relied on comments from a single social media platform, such as using the sentiment of comments [27–30] or generated comments [31,32]. Additionally, the literature has not utilized comments for cross-domain analysis. This limitation neglects the potential insights that can be drawn from analyzing user comments originating from multiple platforms. Incorporating comments from multiple social media platforms into cross-domain fake news detection is important due to the unique characteristics, user behaviors, and linguistic nuances inherent to each platform. Social media platforms cater to diverse user demographics, fostering distinct patterns of interaction and engagement. These variations in communication styles provide complementary insights into the ways fake news is perceived, debated, and propagated across different user communities. Relying solely on a single platform limits the scope of analysis and introduces the risk of bias, as each platform's user base and interaction dynamics may not fully represent broader public discourse. These aspects are notably absent in the existing literature. This gap in the research raises an important question: is it possible to enhance cross-domain fake news detection by integrating user comments from multiple social media platforms? Furthermore, if such integration is feasible, a subsequent question emerges: how can these diverse comments be effectively fused to improve the accuracy of fake news identification? Another important research question is do we rely on comments solely or do we also need to depend on the news content or need to trade off when we need either method?

To solve these questions, we propose a combined approach of contents and comments to detect fake news in a better way. Our approach involves gathering information such as comments from various social media platforms instead of using a single one in order to capture correlations among comments across platforms. Comments are important elements on these platforms that reveal public opinion about news beyond the domain. So, analyzing comments gives more generalization of domain-making detection of fake news across the domain. In the literature, there are some works that use comments by social media users on news for identifying fake news. However, they do not analyze news content and comments separately in terms of authenticity, nor utilize correlation between them. Also, they consider neither the relationship between multiple social media comments nor the way true and fake news travels on social media in terms of comments. Our experiments reveal that true news is rich in content, whereas fake news is rich in comments. Therefore, we incorporate all these elements together in our proposed evidenced fused cross-domain (EFCD) fake news detection method. By bringing content and evidence from different platforms together, our method aims to provide a more complete and precise understanding of how fake news spreads across these platforms. We also observe that not all comments carry the same level of reliability or credibility, indicating some degree of uncertainly associated with comments. By incorporating uncertainty in our analysis, fake news detection accuracy can be further improved.

Furthermore, there is a growing body of research investigating the biases inherent in LLMs, which can perpetuate gender, racial, cultural, and political stereotypes [33–35], ultimately impacting public perception and decision-making processes, particularly in the context of fake news, which has significant effects on society, especially in politics [5–7,26]. These biases pose significant challenges, particularly in critical areas such as fake news detection, where LLMs may inadvertently reinforce fake news rather than mitigate it. To address these concerns, our approach incorporates mechanisms aimed at reducing these biases. Specifically, we utilize comments from diverse news sources to introduce a broader range of perspectives, thereby counteracting the influence of any singular biased viewpoint. Additionally, we introduce an uncertainty-based fusion method that integrates information from multiple sources by consistency checking using correlation, enhancing the balance and fairness of the outputs. This approach not only strengthens the reliability of LLMs in detecting fake news but also contributes to the broader societal discourse on the ethical and responsible use of AI. In summary, the main contributions of this article are:

1. A Dempster–Shafer-based fusion method is proposed to combine probabilities from comment-based detectors across multiple social media platforms, incorporating correlation-based consistency to handle uncertainty effectively.
2. To trade off between content- and comment-based methods, we join them using a threshold that improves the overall fake news identification accuracy rate by 7% compared with previously proposed methods in the cross domain.

2. Literature Review

Some of the main works related to fake news detection research will be presented below. We divide our literature into two parts: (a) in-domain, which includes (i) content-based and (ii) social media-based research, and (b) cross-domain research.

2.1. In-Domain Fake News Detection

Research on fake news detection primarily focuses on in-domain literature. In the context of this research, 'in-domain' refers to situations where both the training and test datasets belong to the same domain. Some examples of domain include war, COVID-19, vaccines, etc. Research on fake news detection primarily focuses on in-domain literature, indicating a significant emphasis on methodologies and technologies developed and evaluated within the sphere of specific content domains. In the context of this research, 'in-domain' is a term that specifically refers to situations where both the training and testing datasets belong to the same domain or subject area. This approach ensures that the models developed for fake news detection are finely tuned to the nuances and characteristics typical of the content within that domain. The rationale behind this focus stems from the understanding that fake news often contains subtle cues and context-specific indicators that are best identified when models are trained and tested on data from the same domain. This enhances the accuracy of detection algorithms, enabling them to effectively discern between true and fake news within tightly defined content boundaries.

- Content-based FND: content-based methods (news text) are the most commonly used techniques for fake news detection in in-domain scenarios. These methods are currently focused on utilizing transformer-based models for feature extraction and classification. Kaliyar et al. [36] used over 8 million tweets about the U.S. general election to develop a bidirectional training approach. This method improves fake news classification by capturing semantic and long-distance dependencies, achieving a 98.90% accuracy with a BERT-based model. Ahn et al. [37] fine-tuned BERT for detecting fake news in a Korean dataset, achieving an ROC-AUC score of 83.8%. Safaya et al. [38] proposed a BERT-CNN model, which outperformed five state-of-the-art models in F1-score on Arabic, Greek, and Turkish tweets, suggesting potential improvements for other languages. In addition to transformer-based methods, TF-IDF, part-of-speech tagging, and word embeddings are also common in content-based fake news detection [39–41]. Since BERT is a highly powerful semantic feature extractor, it is widely used in numerous studies [42,43]. However, its computational intensity can be a limiting factor in some applications. In addition to BERT-based models, other transformers like RoBERT, XBert, and GPT perform well for fake news detection [44,45]. He et al. [46] introduced a single-layer CNN model integrated with BERT, evaluated on the Airline Travel Information Systems (ATIS) dataset, achieving 98.54% accuracy. They noted the model's suitability for short sentences and potential limitations in robustness. Other attention- and sentiment-based methods have also been studied in the literature [30,47,48]. However, these methods may struggle with complex language and context variability.
- Social media- based FND: detecting fake news in traditional news media primarily relies on the content of the news itself. However, in social media, additional contextual information such as user profiles, comments, and news propagation patterns can assist in identifying fake news. The role of users is crucial in this context, as both humans and bots can disseminate news. Users provide valuable information for fake news detection, and user-based features reflect the characteristics of those interacting with news on social media. These features are categorized into individual and group levels.

Individual-level features assess each user's credibility and reliability by examining demographics such as account age, number of followers/following, and the volume of tweets authored [49]. On the group level, it is assumed that the communities spreading fake news differ from those spreading real news. Group-level features are typically derived by aggregating individual-level features, such as the percentage of verified users and the average number of followers within a group [50].

Another important feature in social media is user comments. Several approaches have used temporal linguistic features extracted from sequences of user comments for FND. These methods often rely on fusion techniques that incorporate both news article content and user comments to enhance classification accuracy. In early fusion, features from the text and comments are concatenated in the initial stage, allowing the model to learn joint representations, which can be useful when the modalities are complementary. Late fusion processes the text and comments separately and combines the outputs later, typically by averaging or weighting predictions. This approach ensures that each modality retains its distinctive characteristics. (Early fusion combines data from multiple modalities at an initial stage, allowing the model to capture complex interactions between modalities, such as text and images, from the beginning. However, this approach may lead to overfitting when the interactions between modalities are weak or nonexistent. Late fusion, on the other hand, processes each modality independently and combines their outputs at a later stage, preserving the uniqueness of each modality and reducing the risk of overfitting. However, it may miss valuable cross-modal dependencies by treating the modalities separately for most of the process. Hybrid fusion combines elements of both early and late fusion, capturing some cross-modal interactions early while preserving the distinct information of each modality later on. While powerful, hybrid fusion can increase model complexity and computational demands.) These methods, which fuse text and comments, enhance detection accuracy more compared with relying solely on text. For instance, Ma et al. [51] used recurrent neural networks (RNNs) to analyse sequences of user comments using gated recurrent units, achieving an improvement in accuracy of approximately 15% with gated recurrent units (GRUs) compared with traditional machine learning ML models. Theirs method was further refined by Zubiaga et al. [52], who categorized classified user comments into various categories such as help, reject, question, and comment. They emphasized that the nature of user responses differs depending on the dissemination phase of the news. Similarly, Qian et al. [53] found that fake news tends to elicits more negative responses and questions compared with real news, particularly in the early stages when users have difficulty assessing its credibility. Recent studies have increasingly focused on the sentiment and emotion analysis of comments as a means to enhance the accuracy of fake news detection models. For instance, Guo et al. [28] reported an emotion-based framework that incorporates both publisher details and social emotions that improves the detection accuracy to a notable 87% by considering emotional signals from the content and user comments. In 2023, Hamed et al. [27] demonstrated the importance of using sentiment and emotion analysis in FND. Their approach achieved 90% accuracy, although they acknowledged the complexity of accurately capturing and interpreting emotional features from diverse social media data.

Despite these advancements, explainability remains challenging. Many studies, such as Shu et al. [54] and Sharma et al. [55], offer explanations for their predictions on the basis of particular sentences and user comments. Shu et al.'s model achieved 90% accuracy, and Sharma et al.'s method saw a 2% improvement in accuracy over the previous approach [54]. However, the reliance on high-quality labeled datasets for training makes it difficult to generalize these models across different domains. Only a few systems possess robust explainability features, which are critical for gaining trust and ensuring that the models' predictions are transparent to users. Additionally, the concept of generated comments, as proposed by Nan et al. [32] in 2024, introduces a novel approach to fake news detection. By leveraging large language models (LLMs) to generate diverse comments, this method aims to enrich the dataset and capture a broader range of user interactions. Nan et al.'s model achieved an accuracy of 89%, but the effectiveness of this approach depends heavily on the quality and representation of the generated comments.

Although sentiment and emotion analysis are becoming integral to FND, the lack of explainability in many systems poses a barrier to their widespread adoption. The introduction of generated comments represents a promising direction, but further exploration is required to ensure that they can reliably enhance detection capability across different social media platforms. A limitation of comment-based FND is that it is difficult to locate unbiased comments among the rich information on social media. Because each social media platform is distinct, the same fake news on different platforms will have unique users and unique comments. To address this problem, leveraging comments from multiple social media platforms could be a potential solution. Ensuring consistency among comments across platforms can lead to improved detection accuracy. Comments from different platforms capture varied user viewpoints, actions, and language styles, offering complementary insights that can greatly improve fake news detection. Evaluating the consistency of comments across platforms can further improve detection reliability and accuracy. However, one of the primary reasons this approach remains unexplored in the literature is the lack of available datasets containing comments on the same news across multiple social media platforms. Collecting datasets from social media platforms is time-consuming and they often have API restrictions. This absence of multi-platform datasets highlights a critical gap, underscoring the need for research efforts to develop and utilize such datasets to advance cross-platform fake news detection methodologies.

2.2. Cross-Domain Fake News Detection

Cross-domain fake news detection aims to develop models capable of predicting and identifying future instances of fake news that may not belong to the same domains as those used during the model's training phase. The primary goal of such models is to anticipate and adapt to previously unseen types of fake news, making them highly versatile and robust. However, this area of research encounters significant challenges due to its complexity and the dynamic nature of fake news, which continually evolves across different domains. These challenges include:

1. Lack of diverse data: finding comprehensive datasets that span multiple domains represents a substantial challenge. The diversity and breadth of such data are crucial for building models that can accurately detect fake news across a wide range of topics and formats. Without extensive and varied datasets, models may struggle to generalize well beyond their training environments.
2. Domain-specific features: adapting features that are specific to one domain for use in another is inherently difficult. Features that are highly indicative of fake news in one context may not be relevant or could even be misleading in another. This necessitates the development of sophisticated algorithms capable of identifying and leveraging transferable features that maintain their significance across various domains.
3. Limited transfer learning techniques: although transfer learning offers a promising approach to cross-domain fake news detection, existing methods are still under

refinement. Developing techniques that can seamlessly transfer knowledge from one domain to another without significant loss of accuracy or relevance remains a key research focus. Enhancements in this area are critical for creating models that can adapt to new and emerging forms of misinformation with minimal need for retraining.

In the literature, cross-domain fake news detection primarily concentrates on the structure of the news. The underlying concept is that, regardless of the domain, all news shares a common structure, which can include aspects such as the number of parts of speech, punctuation symbols, and others. Perez et al. [22] were among the pioneers in analyzing fake news detection (FND) using a cross-domain approach. They demonstrated that their proposed method performed well within the same domain, achieving an accuracy of 74%, but its effectiveness significantly decreased to 56% in cross-domain analysis. Additionally, they introduced two datasets: FakeNewsAMT and Celebrity.

Building on this foundation, in 2020, Gautam et al. [23] proposed another cross-domain analysis method utilizing tools such as Spinbot (for paraphrasing), Grammarly (for grammar checking), and GloVe (for word embedding). While their method achieved a 95% accuracy on the FakeNewsAMT dataset, the accuracy dropped to 70% when the model was trained on the Celebrity dataset. This highlighted the ongoing challenge of maintaining high performance across different domains.

Tanik et al. [24] further explored this issue by introducing an ELMo-based word embedding approach. Their method showed an in-domain accuracy of 83.3% but a reduced cross-domain accuracy of 68.5%, indicating a decline from previous methods [22,23]. This trend underscores the difficulty of achieving robust cross-domain performance in fake news detection. Continuing this line of research, Goel et al. [25] experimented with RoBERT, a sentence embedding-based transformer model. Although it exhibited excellent in-domain performance with 99% accuracy on the FakeNewsAMT dataset, it faced significant performance degradation in the cross-domain scenario, with accuracy dropping to 59% when trained on the Celebrity dataset. Finally, Jannatul et al. [56] used structured features, including parts of speech with word-based features, and achieved 70% accuracy when tested with the Celebrity dataset. Though the overall accuracy increased, there was a significant loss in accuracy for the fake class.

Thus, various studies in the literature have shown a drastic reduction in performance for cross-domain analysis, and no studies have completely focused on cross-domain issues. Instead, they provide solutions for in-domain problems and show how their method performs for cross-domain scenarios. Therefore, we are motivated to design a solution that exclusively focuses on cross-domain fake news detection.

3. System Architecture

The proposed evidence fused cross-domain (EFCD) fake news detection model is composed of three main modules: the initial screening module (ISM), the social media module (SMM), and the fusion module (FM). The major notations used in this section are listed in Table 1 for reference.

Table 1. List of major notations.

Notation	Description
T	News text being evaluated
p_{base}	Probability of a news text T being fake as determined by the model M_1
T^1	Semantically similar news to T found on Twitter
T^2	Semantically similar news to T found on Reddit
M_1 and M_2	Machine learning models
U	Uncertainty
c_{tw}	Twitter's comments
c_{rd}	Reddit's comments
e_{i_tw}	Encoded comments for the *i-th* comment of Twitter
e_{i_rd}	Encoded comments for the *i-th* comment of Reddit
C	Correlation
$p^c_{i\_tw}$	Probability derived from the *i-th* comment of Twitter by model M_2
$p^c_{i\_rd}$	Probability derived from the *i-th* comment of Reddit by model M_2
$p^{sup}_{i\_tw}$ and $p^{\neg sup}_{i\_tw}$	Support and non-support probability for the *i-th* comment of Reddit
$p^{sup}_{i\_rd}$ and $p^{\neg sup}_{i\_rd}$	Support and non-support probability for the *i-th* comment of Reddit
m_i	Fused probability of *i-th* comment

The ISM employs a machine learning (ML) model to find the probability, denoted as p_{base}, of a news text, with (T) being fake. If p_{base} falls below a predetermined threshold, it is directly accepted as the final probability, hence terminating the further process. However, if p_{base} exceeds this threshold, the system proceeds to the SMM, where it searches and identifies semantically similar news to T, labeled T^1 and T^2, on two different social media platforms, respectively. (Since each social media platform is unique, this article specifically focuses on Twitter and Reddit. Therefore, the method proposed here is designed based on these two platforms. When we refer to multiple social media platforms, we specifically mean Twitter and Reddit.) Comments associated with T^1 and T^2 are collected, and their correlation is assessed by forming pairs based on the chronological order of their timestamps. An uncertainty measure is calculated to support the possibility of the news being fake from the correlation. In the FM, probabilities from all the comments are fused using Dempster–Shafer theory (DST), resulting in a fused probability. Ultimately, the system aggregates these fused probabilities to derive a final probability, which serves as the final probability of T being fake. The entire process is illustrated in Figure 1. The detailed description is gives below.

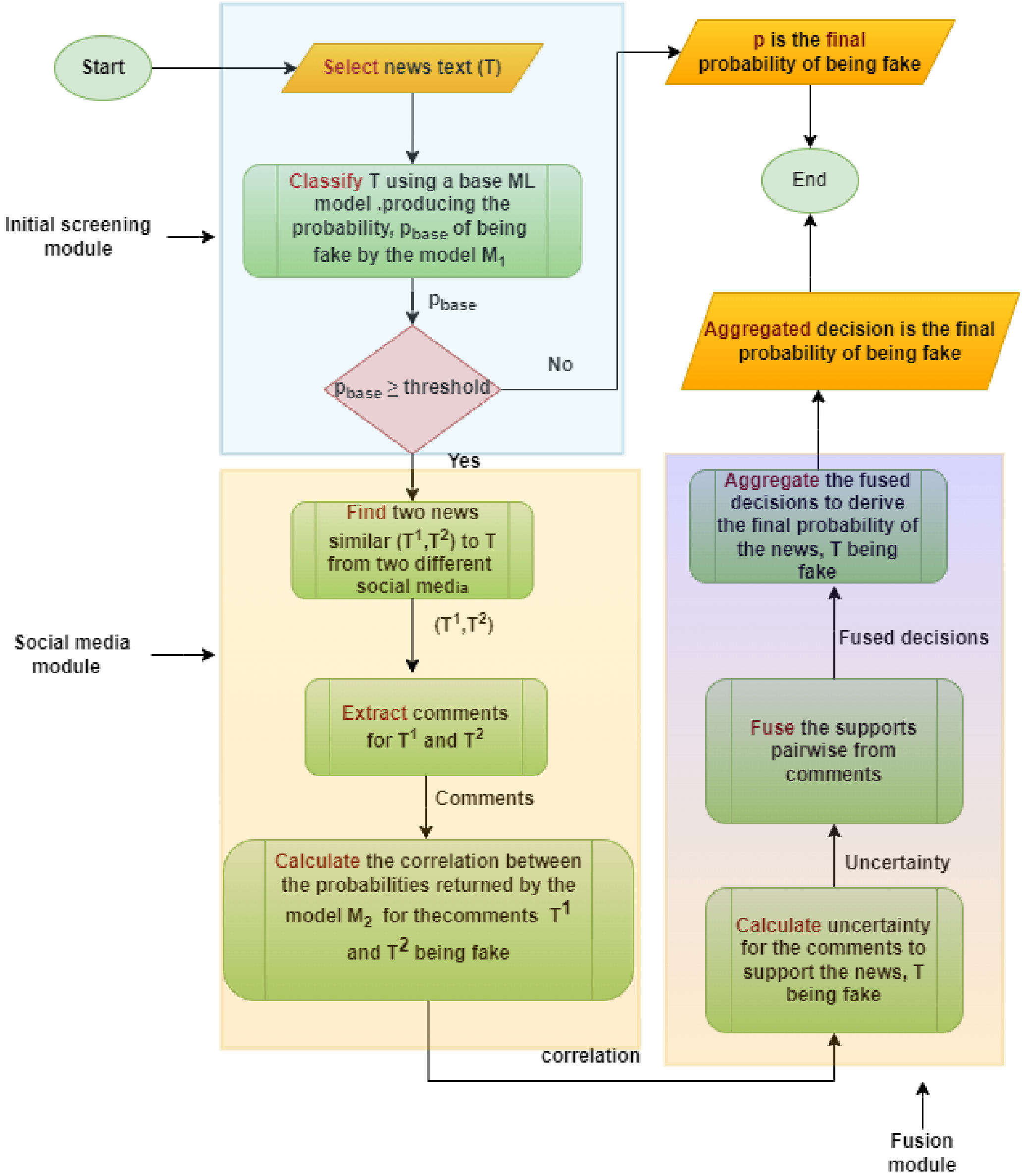

Figure 1. System overview for the proposed EFCD method.

1. Initial screening module: Figure 1 describes the initial screening process where the probability, p, is calculated using linear regression (LR) classified model, which we call M_1. This classifier employs an extensive feature set crafted by merging two main types of features: (i) features generated using part-of-speech (POS) tags; and (ii) features derived directly from the text, emphasizing word-level characteristics. POS tagging can often be ambiguous, and the introduction of word tags aims to mitigate this ambiguity [56]. A list of features used in training the LR classifier, M_1, is shown in Table 2. New news text, T, is input into the trained model, M_1, to output p_{base}. Mathematically, it can be written as:

$$p_{base} \leftarrow M_1(T)$$

After the initial screening by this method the value of p_{base} is compared against a threshold (Section 5 sheds more light on how to determine the threshold value), and if the value of p_{base} is lower than the threshold then the news is deemed be true with high confidence and the process is terminated. Otherwise, the process moves to the next social media module (SMM).

Table 2. Features used in training the classifier model in the ISM module [56].

Word tags	Word count, char count, average word count, hashtags count, link count, number of length, user mention count
Pos tags	CC—Coordinating conjunction, CD—Cardinal number, DT—Determiner, EX—Existential there, FW—Foreign word, IN-Preposition, JJ—Simple adjective, JJR—Comparative adjective, JJS—Superlative adjective, MD—Modal, NN—Singular noun, NNP—Noun proper singular, NNPS—Noun proper plural, NNS—Noun plural, PDT—Pre-determiner, POS—Possessive ending, PRP—Personal pronoun, PRP$—Possessive pronoun, RB—Adverb, RBR—Adverb comparative, RBS—Adverb superlative, RP—Particle, SYM—Symbol, TO—to, UH—Interjection, VB—Verb base form, VBD—Verb past form, VBG—Verb present or gerund particle, VBN—Verb past participle, VBP—Verb 3rd person singular, LS—List marker, VBZ—Verb 3rd person singular, WDT, WP—Wh determiner, WP$—possessive wh pronoun, WRB—Wh adverb and other symbols

2. Social media module: in the social media module, we collect similar news from the two different social media platforms and also the comments. Subsequently, we compute the probability of the comments being fake with respect to the news, T. The details are discussed below:
    - Collection of similar news articles: here we consider collecting news items similar to T from two online social networks (OSNs). Collecting similar news articles from various OSNs, such as Twitter and Reddit, is an important step in our fake news detection system. Each platform attracts a unique user base with diverse content, opinions, and sources. By gathering news articles from multiple platforms, we ensure a comprehensive coverage of the news landscape, capturing different perspectives and reducing the risk of bias. The detailed procedure of collecting similar news from social media is described in Section 4 (Dataset Preparation).
    In addition to news articles, we collect the comments associated with these articles from social media platforms, because comments offer valuable insights into public sentiment, reactions, and potential biases related to the news items [57–60]. Analyzing these comments provides a deeper understanding of how users perceive and respond to the news. By considering this user-generated content, we can evaluate the overall credibility and reception of the news among the online community.
    - Deriving probability from comments: after collecting comments, the probability of news, T, being fake is calculated by Algorithm 1, named as 'Fake news detection from comments (FNDC)'. Some important aspects of this algorithm are given below:
        – FNDC analyzes the content and characteristics of each comment using natural language processing (NLP) and machine learning techniques. It examines the language patterns of the comment using BERT sentence embeddings to identify potential indicators of fake information [61]. The 768 features of BERT provide rich contextual representations and fine-grained semantic understanding at the sentence level, allowing FNDC to capture the nuanced meaning of each comment. For example, during the COVID-19 pandemic, comments like 'This miracle herb can cure COVID-19 overnight, but doctors are keeping it a secret!' were flagged. BERT's embeddings helped in detecting sensationalist language, such as the adjectives 'miracle' and 'overnight' [56,62]. Similarly, in the context of the Russia–Ukraine war, state-

ments like 'Ukrainian forces have all surrendered, and this is being hidden by Western media!' were identified. The use of absolute terms like 'all', combined with verbs implying secrecy such as 'hidden', were key indicators [63,64]. By leveraging these detailed BERT features, FNDC effectively enhances the model's ability to identify and mitigate unreliable information, ensuring the integrity of the data analyzed. We call this FNDC model M_2.

– M_2 (i.e., FNDC) operates under the assumption that, if news is fake, the comments on the corresponding news article are also likely to be fake. Similarly, if news is true, it suggests that the associated comments are more likely to be true.

The detailed algorithm is given below, where input is a set of n comments, c_n, and the probability of these nti-th comments being fake is denoted as p_n^c:

Algorithm 1 Fake news detection from comments (FNDC)

Require: List of comments from social media: c_n
Ensure: List of probabilities of comments being fake: p_n^c
Preprocessing: Clean and preprocess the comments to get $clean_{cn}$
Use Nearmiss algorithm for balancing data
NLP Analysis: Extract features from c_n^{clean} using BERT, which is called encoded comments.
Train Classifier: Use labeled data to train an MLP with $features$ as input to get model M_2
$p_{cn} \leftarrow$ Empty List
for each $feature$ in $features$ **do**
    $prob \leftarrow$ MLP.predict($feature$)
    Append $prob$ to p_n^c
end for
Return p_n^c

So, for the Twitter comments, c_{tw}, the probabilities of them being fake are denoted as $p_{i\_tw}^c$, and for Reddit comments, they are denoted as $p_{i\_rd}^c$, where both i_tw and i_rd range from 1 to n. The calculation of n is as follows: if there are c_{tw} comments from Twitter and c_{rd} comments from Reddit, the number of comments considered in the analysis is:

$$n = \min(c_{tw}, c_{rd})$$

- Fusion module: from the last module, we find the probability of news being fake with respect to each comment. Now, we want to fuse the probability with uncertainty. The fusion module consists of three parts: correlation calculation, uncertainty calculation, and aggregation of the fused decision. The correlation calculation aims to understand the relationship between comments across platforms, while the uncertainty calculation assesses the reliability of the comments. Finally, the aggregation step combines these analyses to arrive at a unified decision about the credibility of the news.

  – Correlation calculation: to better understand the relationships between comments on two social media platforms, we calculated the correlations between them. Specifically, we measured the correlation between the probabilities of news being fake based on comments from each platform, using the FNDC module. This statistical measure helps quantify how closely the fake news probabilities from one platform align with those from the other. The correlation coefficient, C, ranges from -1 to 1. Positive values ($C > 0$)

indicate that, as the probability of fake news increases on one platform, it tends to increase on the other platform as well. This suggests a similar trend in fake news probabilities across both platforms. Conversely, negative values ($C < 0$) suggest an inverse relationship, where a higher probability of fake news on one platform corresponds to a lower probability on the other. Understanding these correlations helps us analyze how social media platforms interact and influence the spread of fake news. Table 3 illustrates our correlation calculation with an example involving comments from Twitter and Reddit about a news item detailing a confrontation between Roger Federer and Frances Tiafoe. Each comment of each platform (Twitter and Reddit) was encoded using BERT into a 768-dimensional vector. Let e_{i-tw} and e_{i-rd} denote the encoded comment for the i-th comment on Twitter and Reddit, respectively. These encoded comments were then processed through a machine learning model, M_2, to calculate a probability for each comment. Mathematically, this is expressed generally for Twitter and Reddit as:

$$p^c_{i\_tw} \leftarrow M_2(e_{i\_tw})$$

and

$$p^c_{i\_rd} \leftarrow M_2(e_{i\_rd})$$

where $p^c_{i\_tw}$ and $p^c_{i\_rd}$ are the probability for the i-th comment on Twitter and Reddit, respectively.
In Table 3, the original comments, their encoded vectors, and the computed probabilities are presented for both Twitter and Reddit. For instance, a Twitter comment 'Doesn't account schedule Federer chooses play…' is encoded and the corresponding probability is 0.08, while a Reddit comment 'It 100 the time I topic /r/tennis' is encoded and has a probability of 0.91.
After calculating all probabilities using Algorithm 1, the correlation between the probabilities is calculated. To calculate these probabilities, pairs are formed between Twitter and Reddit comments based on their timestamps. This means the first comment on Twitter is paired with the first comment on Reddit, and so on. Specifically, the Pearson correlation coefficient (C) is determined to evaluate the relationship between the sets of probabilities.

$$\{p^c_{i\_tw}\} \quad \text{and} \quad \{p^c_{i\_rd}\}$$

The correlation value (Table 3) of -0.61 reveals a moderate-to-strong inverse relationship between comments on Twitter and Reddit, highlighting contrasting patterns of user engagement and interpretation across the two platforms. When comments on Twitter align positively with a news item, Reddit comments often challenge or refute it, and vice versa. This divergence reflects the distinct user behaviors and discussion dynamics unique to each platform—Twitter's real-time and concise communication contrasts with Reddit's preference for detailed and critical discussions. This negative correlation underscores the variability in how users interact with the same news item across platforms, shaped by their respective functional and cultural characteristics. Such findings reinforce the importance of cross-platform analysis in understanding how information is debated and interpreted, offering valuable insights into the mechanisms of news dissemination and the potential for misinformation to spread.

Table 3. Calculation of correlation values from the comments of multiple platforms.

News item:	Roger Federer has fist fight with Frances Tiafoe after Miami Open defeat. Things became extremely heated between world class tennis star Roger Federer and American teenager and tennis up and comer Frances Tiafoe after their Miami Open tennis match. The match was not close, but the two players were playing in rainy and windy conditions, which gave Federer an edge with his years of experience over Tiafoe. After Federer beat the younger tennis professional in three sets, the two players began to yell at each other. Tiafoe was angry about several alleged incorrect calls made by Federer in the match. Tiafoe then jumped over the net and attacked Federer with several punches. Federer defended himself until several observers came and broke up the fight. The two have both issued public apologies to their fans and to each other, but clearly things will not be settled until they face each other on the court another time.					
Twitter			**Reddit**			**Correlation, C**
Comments, c_{tw}	**Encoded comments, e_{tw}**	**Probability, $p^c_{i-tw} \leftarrow M_2(e_i)$**	**Comments, c_{rd}**	**Encoded comments, e_{rd}**	**Probability, $p^c_{i-rd} \leftarrow M_2(e_i)$**	
Doesn't account schedule Federer chooses play. I mean didn't bypass French prioritise grasscourt season.	[−0.51961124, 0.36066872, 1.1010087, …]	0.08	It 100th time I topic /r/tennis	$[-5.89307308 \times 10^{-1}$, $6.89936399 \times 10^{-1}$, $5.71980417 \times 10^{-1}$, …]	0.91	
Managed clinch important must-win match fired isner. respect man's name	$[-7.77983904 \times 10^{-1}$, $2.72371382 \times 10^{-1}$, $7.28409469 \times 10^{-1}$, …]	0.99	Nice. don't want contribute I think add categorie "Mental Strenght" seeing Federer "Saving BPs" horrendous BP conversion rate doesn't fit.	$[-5.32212555 \times 10^{-1}$, $3.82354587 \times 10^{-1}$, $1.10202396 \times 10^{+0}$, …]	0.03	**−0.61**
Goat here, shit man	$[5.41974604 \times 10^{-1}$, $7.15636134 \times 10^{-1}$, $2.83419457 \times 10^{-3}$, …]	0.99	Needs Nadals bald patch	$[1.14744902 \times 10^{-1}$, $8.40671659 \times 10^{-1}$, $6.28644377 \times 10^{-2}$, …]	0.99	
.	.	.	.	.	.	
.	.	.	.	.	.	
.	.	.	.	.	.	

– Comment analysis with uncertainty: after calculating the correlation, the next step is uncertainty. This process is divided into three parts. First, we calculate the uncertainty based on the correlation between comments. Next, we compute both the support and non-support probabilities to quantify the likelihood of the news being fake or true in the presence of uncertainty. Finally, we fuse these decisions to arrive at a comprehensive conclusion.

  ◦ Uncertainty calculation: analyzing comments from two social media platforms can introduce the possibility of uncertainty that needs to be addressed. Calculating the uncertainty associated with comment analysis provides a measure of the confidence or reliability of the analysis results. This information aids in interpreting the findings and making informed decisions or drawing accurate conclusions based on the analyzed comments with the presence of uncertainty. Here we measure uncertainty, U, as defined by the following equation:

$$U = \begin{cases} 1, & \text{if } C = 0 \\ \frac{(1-|C|)\cdot\log(|C|,10)\cdot\log(|C|,10)}{\log(0.001,10)}, & otherwise \end{cases} \quad (1)$$

Figure 2 illustrates the relationship between correlation (C) and uncertainty (U), and comes from Equation (1). When there is a positive or negative correlation, the level of uncertainty is relatively low. However, as C approaches zero, U increases significantly. When $C = 0$, the uncertainty reaches its maximum value of 1.

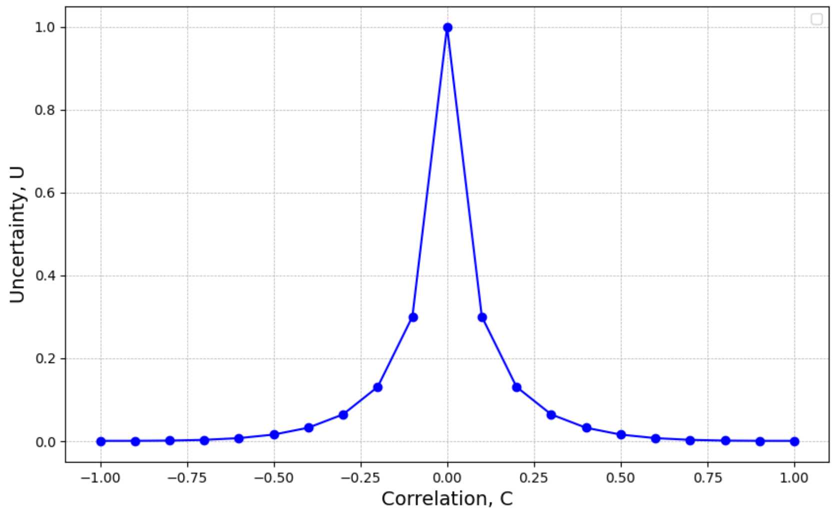

Figure 2. Relationship between uncertainty and correlation.

The justification of this equation for uncertainty, U, is linked to Shannon's information theory, which quantifies uncertainty as a measure of unpredictability in probabilistic systems. When the correlation (C) between two social media platforms is zero ($C = 0$), the platforms exhibit no relationship, leading to maximum uncertainty ($U = 1$), consistent with the concept of maximum entropy in the absence of information. As the absolute correlation ($|C|$) increases, uncertainty decreases, reflecting stronger consistency in evidence across platforms. The term $1 - |C|$ models this reduction linearly, while the logarithmic components $\log(|C|, 10) \cdot \log(|C|, 10)$ amplify the sensitivity to weak correlations ($|C| \to 0$), where the lack of agreement leads to higher uncertainty. The normalization factor $\log(0.001, 10)$ ensures U remains bounded and interpretable. This formulation is an adaptation of entropy-based principles commonly applied in communication systems [65], making it suitable for modeling uncertainty in cross-platform fake news detection.

We have used the same uncertainty calculation method across different social media platforms for all comments to ensure fairness and comparability in the analysis.

- Support and non-support probability calculation: assessing the level of support, p^{sup}, or non-support, $p^{\neg sup}$, expressed in comments is another key factor for evaluating the credibility of information or detecting fake news. Support probability means the probability of news being fake in the presence of uncertainty and the inverse for non-support. For any comment, the general rule is this:

$$p^{sup} + p^{\neg sup} + U = 1 \tag{2}$$

So, the calculation of support probability for Twitter follows as:

$$p^{\text{sup}}_{i-tw} = \begin{cases} 0, & \text{when } U \approx 1 \\ \min\left(\left\{p^{c}_{i\_tw}\right\}^{n}_{i\_tw=1}, 1 - U\right), & \text{otherwise} \end{cases} \tag{3}$$

Similarly, the support probability for Reddit is defined as:

$$p_{i\_rd}^{\text{sup}} = \begin{cases} 0, & \text{when } U \approx 1 \\ \min\left(\left\{p_{i\_rd}^{c}\right\}_{i\_rd=1}^{n}, 1 - U\right), & \text{otherwise} \end{cases} \tag{4}$$

So, after completing this step, we obtain a 'support' probability and a 'non-support' probability for each comment on both social media platforms. Given that we have n comments from both Twitter and Reddit, after this step we have n support and n non-support probabilities for each platform. Additionally, a common uncertainty, U, value is associated with every comment across all platforms.

- Fuse the support probabilities: afterwards, the pairwise support probabilities of comments from both Twitter and Reddit are fused using the Dempster–Shafer (DS) combination rule. This approach involves combining the support probabilities from both platforms according to the equation below.

  The combined support probability for the i-th comment using Dempster–Shafer theory [66] is:

$$m_i = \frac{p_{i\_tw}^{\text{sup}} \cdot p_{i\_rd}^{\text{sup}} + p_{i\_tw}^{\text{sup}} \cdot U + U \cdot p_{i\_rd}^{\text{sup}}}{1 - (p_{i\_tw}^{\text{sup}} \cdot p_{i\_rd}^{\neg\text{sup}} + p_{i\_tw}^{\neg\text{sup}} \cdot p_{i\_rd}^{\text{sup}})} \tag{5}$$

  Since there are a total of n comments, the combined values form a list, m_i, for $i = 1, 2, \ldots, n$.

– Aggregate the fused decision: after fusing the decisions from comments on two different platforms using Dempster–Shafer theory, we aggregate the combined support probabilities, m_i, to obtain the final probability, p_{final}. This is done by first determining the majority class of the support values (whether more values are equal or greater than 0.5 or less than 0.5) and then averaging the support values that belong to the majority class. This approach ensures a comprehensive and balanced assessment of the information gathered from both Twitter and Reddit comments.

$$P_{\text{final}} = \frac{1}{g} \sum_{i \in \text{majority}} m_i \tag{6}$$

where g is the number of m_i values in the majority class.

– Hypothetical case study: COVID-19 vaccine misinformation: to illustrate the applicability of our model, consider a hypothetical fake news case claiming that "alcohol cures COVID-19". Our proposed system would first analyze the content of the news article. If the content is rich, detailed, and linguistically credible, the system assigns a low probability of the news being fake, indicating it is likely true. However, if the content is shallow or lacks depth, the system assigns a high probability of the news being fake. In such cases, the model further analyzes user comments from multiple platforms, such as Twitter and Reddit, to gain additional insights.

  For this example, comments might display varying degrees of support or disagreement. Supportive comments might include:

  (a) "Finally, an easy cure for COVID! Alcohol every day is the way!"
  (b) "This is amazing! People need to know alcohol can save lives!"

Conversely, non-supportive comments might include:

(a) "This is fake news. Drinking alcohol won't cure COVID-19."

(b) "Don't trust this claim—it's dangerous and unsupported by science!"

If comments across both platforms consistently support or refute the claim, this consistency strengthens the evidence, enhancing the model's confidence in its prediction. For instance, consistent disagreement across platforms would reinforce the conclusion that the news is fake with lower uncertainty. On the other hand, if comments display conflicting patterns—supportive on one platform but refuting on another—the model incorporates high uncertainty. Correlation is calculated to check the consistency between comments of two social media platforms. By integrating content analysis and cross-platform comment evaluations, our system effectively detects fake news while addressing the nuances of cross-domain fake news. This case highlights the robustness of our approach in leveraging both content and user interactions to assess the credibility of news, even in scenarios involving conflicting or ambiguous signals.

4. Dataset Preparation

To train and evaluate our model in the context of cross-domain fake news detection, we utilized two distinct datasets: training was conducted using the 'Celebrity' dataset [22], and testing was performed using the dataset 'FakeNewsAMT' [22]. This was done so that our model that was built in one domain could be tested in another domain. Other studies [22–24] also adopted similar datasets to evaluate their proposed methods. The Celebrity dataset consists of 500 news articles pertaining to celebrity gossip. The articles were gathered from entertainment-focused online magazines, such as Entertainment Weekly, People Magazine, and Radar Online. This dataset includes an equal number of fake and true news items, enabling balanced training and testing for FND models. On the other hand, the FakeNewsAMT dataset comprises 480 news articles evenly split between fake and genuine news. The genuine news was sourced from various mainstream American news websites, including those of ABC News, Cable News Network, USA Today, The New York Times, Fox News, Bloomberg, and CNET. Fake news was created through crowdsourcing using Amazon Mechanical Turk. Workers were instructed to produce fake versions of genuine news articles while preserving the same topic and length and avoiding implausible content. The dataset spans six domains: technology, education, business, sports, politics, and entertainment, each containing 80 articles equally divided between fake and true. The statistics for the FakeNewsAMT and Celebrity datasets are presented in Table 4.

Table 4. Statistics for the FakeNewsAMT and Celebrity datasets.

Dataset	Total Number of News	Avg Words/Article	Avg Words/Sentence	Distribution (Fake/True)
Celebrity	500	122	24	250/250
FakeNewsAMT	480	132	23	240/240

However, the datasets [22] contained only news content, while our methodology required both news content and associated comments from two different social media platforms. Therefore, we collected our dataset as described below.

Algorithm 2 illustrates the steps for preparing datasets for our proposed method, which focuses on identifying similar news content from social media platforms. We selected Twitter and Reddit as our primary sources for gathering relevant news items and comments.

The objective was to identify the most similar news item on each platform in relation to a given news article, T.

Algorithm 2 Similar news article retrieval from OSNs

Require: News article T, Maximum news items z
Ensure: Similar news articles T^1 from social media
Extract top m keywords from T
$K \leftarrow$ `ExtractKeywords`(T)
Generate prioritized queries
$Q \leftarrow$ `GenerateQueries`(K)
Initialize *results* list
results $\leftarrow$ []
for each query q in Q **do**
    Append `SearchPlatforms`(q) to *results*
    if length(results) $\geq z$ **then**
        break
    end if
end for
Compute BERT features for T
bert_T $\leftarrow$ `ComputeBERTFeatures`(T)
for each news article r in results **do**
    Compute BERT features for r
    bert_r $\leftarrow$ `ComputeBERTFeatures`(r)
    Compute similarity
    $s_r \leftarrow$ `ComputeCosineSimilarity`(bert_T, bert_r)
end for
Sort news articles by similarity in descending order
$T^1 \leftarrow$ First article in sorted results
return T^1

First, we employed the 'ExtractKeywords' function, utilizing the text-rank algorithm, to extract the top m keywords from each article, T. For instance, for a news item titled 'Tech Giants Face New Regulations Over Data Privacy', the extracted keywords might include 'data privacy', 'tech giants', 'regulations', 'user data', and 'transparency'. Using these extracted keywords, we generated prioritized search queries through the 'Generate-Queries(k)' function, which emphasized 'AND' combinations over 'OR' to ensure more precise search results. Example queries included 'data privacy AND tech giants AND regulations' and 'user data AND transparency AND tech giants'. Subsequently, we utilized the 'SearchPlatforms(q)' function to search for related articles on Twitter and Reddit. This process continued until a maximum of z relevant articles were retrieved from each platform. For example, a search query such as 'data privacy AND tech giants AND regulations' might return tweets and Reddit posts discussing how new data privacy laws affect major technology companies. For each extracted article and the original article, T, we computed the BERT features using the 'ComputeBERTFeatures' function. This function leverages the BERT model from Hugging Face's Transformers library to obtain contextualized embeddings for the text data. We then calculated the cosine similarities between the feature vectors of T and each extracted news item using the 'ComputeCosineSimilarity' function from the scikit-learn library. This metric measures the cosine of the angle between two vectors, with a higher cosine similarity score indicating greater similarity. For example, similarity scores might range from 0.88 to 0.95 for different articles.

Based on these cosine similarity scores, we identified the articles most similar to T from Twitter and Reddit. The article with the highest similarity score on Twitter was labeled as T^1, and the article with the highest similarity score on Reddit was labeled as T^2.

5. Experimental Setup

In our study, we utilized the 'Celebrity' dataset, which comprises 480 samples. We partitioned the dataset into two distinct sets for training and validation to thoroughly validate our models. Specifically, 80% of the samples were allocated for training, while the remaining 20% were reserved for validation. Initially, we determined the optimal number of hidden layers and epochs using this configuration. To achieve a robust assessment of the threshold value (theory in Section 3), we conducted experiments leveraging this validation size to enhance the reliability of our evaluation. Additionally, we used the 'FakeNewsAMT' dataset, which also comprises 480 samples, exclusively for testing. Since we aimed to perform cross-domain analysis, we ensured that no data from the 'FakeNewsAMT' dataset were used for training or validation purposes. This approach allowed us to evaluate the model's performance and generalization capability on entirely unseen data. For our experiments, we used accuracy, precision, recall, F1-score, and false negative rate (FNR) as performance metrics, with fake news considered as the positive class. This comprehensive evaluation framework allowed us to thoroughly assess the effectiveness and robustness of our proposed EFCD model.

We employed the Python 3.8 scikit-learn library to develop the models in our experiment. Scikit-learn is a robust machine learning library in Python 3.8 that provides simple and efficient tools for data mining and data [67]. For the first model, M_1, we employed a linear regression (LR) approach. Linear regression is a fundamental statistical technique used to model the relationship between a dependent variable and one or more independent variables by fitting a linear equation to the observed data. In our implementation of the M_1 model, we utilized the default parameters provided by the scikit-learn library. These default parameters include using the ordinary least squares (OLS) method to minimize the sum of the squared residuals between the observed and predicted values. This setup allows for straightforward interpretation and implementation, making it an ideal choice for the initial screening in our study.

For the second model, M_2, we used a deep neural network (DNN), implemented as a Multilayer Perceptron (MLP). This model was fed with the output of 768 features extracted from BERT to produce a binary output: fake or real. We evaluated the performance of the DNN model by computing the mean square error (MSE) and the regression scores (R2-Scores).

Additionally, we utilized the 'Adam' optimizer for MLP, which is widely recognized for its efficacy in weight optimization. The ReLU (rectified linear unit) function was chosen as the activation function for the hidden layers. As depicted in Figure 3, the MSE diminishes as the number of hidden layers increases, reaching its lowest value of 0.1446 at the fifth hidden layer. Conversely, the highest R2-Score, 0.4217, is also achieved at the fifth hidden layer.

We also validated our deep neural network (DNN) model using MSE as the loss value and as a performance indicator across different numbers of epochs with a batch size of 200. The loss values are plotted against epoch numbers in Figure 4. Figure 4 shows that the loss value gradually decreases up to 500 epochs and then increases again. The model obtained the highest precision for 500 epochs with a minimum loss value of 0.1446, which motivated us to use 500 epochs in our experiment.

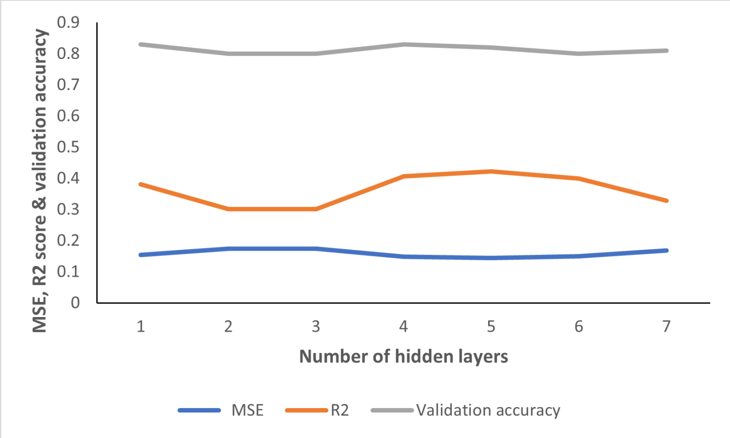

Figure 3. The MSE, R2-Scores, and validation accuracy for different numbers of hidden layers (100 neurons per layer) of DNN for model M_2.

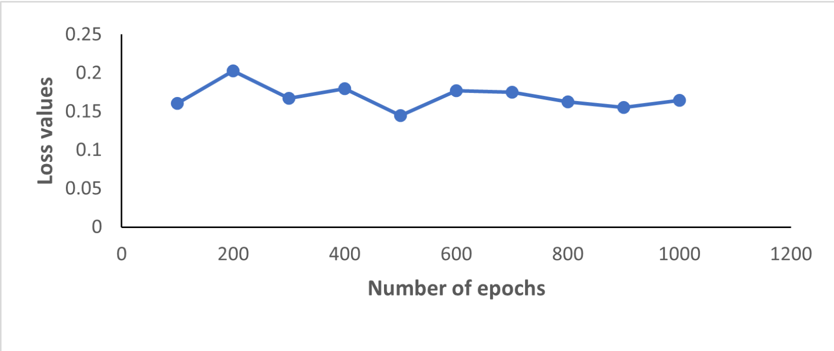

Figure 4. Loss value for the different number of epochs for model M_2.

After selecting the hidden layer, neuron, and epoch, we aimed to set a threshold by which we decided whether to choose content or comment. To determine the best threshold value, we experimented with different thresholds using 0.2 validation size and observed accuracy, precision, and recall, shown in Table 5. We found that decreasing the threshold increases the recall value, which represents our fake news detection accuracy. Conversely, a higher threshold results in increased precision. The highest precision was observed at a threshold of 0.2. The recall value peaked at a threshold of 0.1. In summary, if the primary goal is to maximize fake news detection, a low threshold of 0.1 is recommended. For threshold values of 0.1 and 0.15, recall was high but precision was low, while, for 0.25 and 0.3, accuracy and precision were high but recall was low. However, a threshold of 0.2 is more appropriate to achieve balanced accuracy for both classes because other performance metrics, such as accuracy and precision with recall, were also higher than 0.8. For subsequent experiments, we selected a threshold of 0.2 with a validation size of 0.2.

Table 5. Selection of threshold values for validation size 0.2.

Hidden layer: (100,100,100,100,100) Epoch: 500			
Threshold	**Accuracy**	**Precision**	**Recall**
0.1	0.77	0.73	0.85
0.15	0.82	0.80	0.85
0.2	0.82	0.82	0.81
0.25	0.84	0.88	0.79
0.3	0.84	0.90	0.77

6. Experimental Results

To evaluate the performance of our proposed EFCD model, we conducted three distinct examinations. Initially, we considered only the content of the news, referring to this approach as the 'only content' model. Next, we focused solely on the comments, which we termed 'only comment'. Finally, we assessed the performance of our proposed EFCD model. This comprehensive evaluation allowed us to compare the effectiveness of each approach and highlight the strengths of the EFCD model.

The graph depicted in Figure 5 provides a detailed comparison of the three models designed for the detection of fake news. The 'only content' model, with an accuracy of 0.79, is quite precise in its detection, boasting a high precision of 0.98, which demonstrates it is good at true news detection. However, it has the lowest recall value among these three models, representing the weakness of detecting fake news, which is the main objective of fake news detection. In contrast, the 'only comment' model shows a lower accuracy of 0.53 and a precision of 0.52, which might initially seem less impressive. Nevertheless, it compensates with a high recall of 0.86, indicating its prowess in identifying a wider range of fake news instances. Its F1-score, a balance of precision and recall, stands at 0.65. This model is particularly adept at flagging fake news for review, making it a valuable tool for initial screenings.

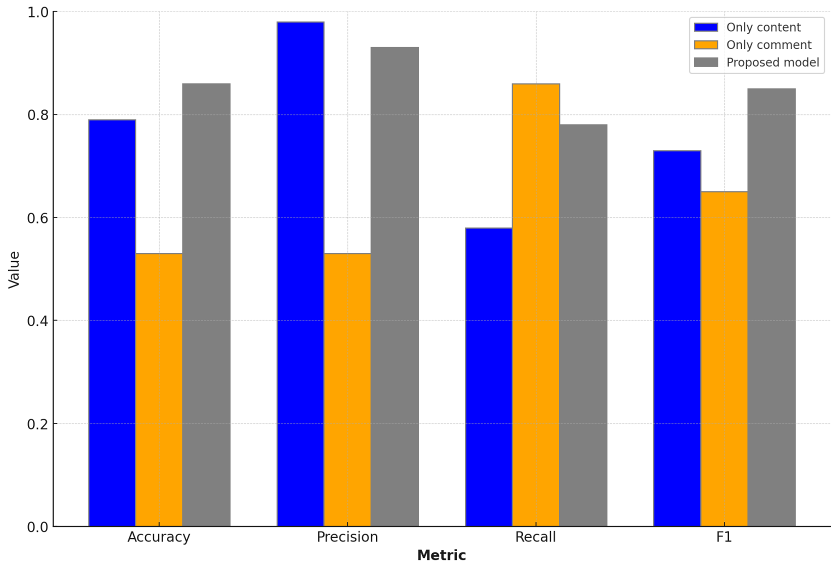

Figure 5. Comparison of accuracy, precision, recall, and F1-score for different methods.

The 'proposed model', which merges the strengths of content-based and comment-based detection by utilizing data from both Twitter and Reddit, excels across all performance metrics. It achieves the highest accuracy of 0.86, a strong precision of 0.93, and a robust recall of 0.78. The aggregate of these metrics is reflected in the highest F1-score of 0.85, confirming the superiority of the combined model approach.

The graph in Figure 6 focuses specifically on the false negative rate (FNR) for the three models. The 'only content' model exhibits a high FNR of 0.42, suggesting that, while it can reliably identify true news, it fails to catch a significant portion of fake news. This is a notable disadvantage for scenarios where the detection of fake news is just as important as the true identification of true news. The 'only comment' model achieves a much lower FNR of 0.14, demonstrating its effectiveness in identifying fake news instances. The 'proposed EFCD model' achieves a balanced FNR of 0.19, showing a substantial improvement over the 'only content' model while not quite matching the performance of the 'only comment' model in terms of FNR.

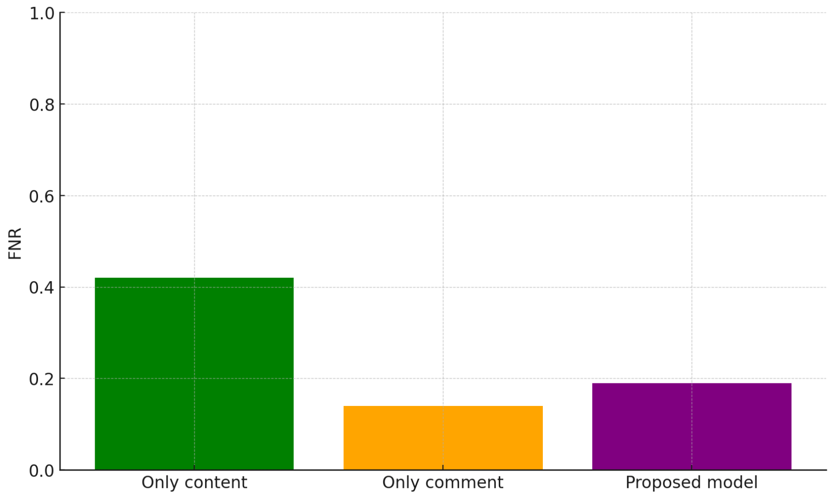

Figure 6. Comparison of false negative rate (FNR) for different models.

Overall, the 'proposed model' (EFCD) stands out as the most reliable system for both identifying true news and flagging fake news. Its comprehensive analysis benefits from the cross-referencing of content and comment indicators, making it an indispensable tool in the modern information landscape where speed and accuracy are paramount. By integrating the strengths of the 'only content' and 'only comment' models, the 'proposed model' (EFCD) offers a significant step forward in the fight against fake news. This evidence-based approach is essential for navigating the complexities of news verification in the digital age.

Comparison with the Previous Methods

In a recent comparative study, featured in Table 6, a variety of models have been evaluated for their effectiveness in detecting fake news across domains. This table lists models developed from 2017 to the present, illustrating their performance in terms of accuracy, recall, precision, F1-score, and false negative rate (FNR).

Table 6. Comparison of the proposed EFCD method with other methods in the current literature.

Training: Celebrity Testing: FakeNewsAMT					
Model	**Accuracy**	**Recall**	**Precision**	**F1-Score**	**FNR**
Perez et al. 2017 [22]	0.64	-	-	-	-
Gautam et al. 2020 [23]	0.7	-	-	-	-
Goel et al. 2021 [25]	0.7	0.59	0.77	0.67	0.40
Jannatul et al. 2022 [56]	0.79	0.58	0.98	0.74	0.42
Proposed	0.86	0.78	0.93	0.85	0.19

Perez et al. [22] introduced an initial model with an accuracy of 0.64; however, recall and FNR metrics were not reported. Subsequently, Gautam et al. [23] advanced the field with a model that achieved an accuracy of 0.7, though recall and FNR values were not disclosed. An incremental improvement was observed with Goel et al. [25], who matched the 0.7 accuracy and reported a recall of 0.59 and an FNR of 0.40, offering a more nuanced understanding of the model's performance. Jannatul et al. [56] further enhanced accuracy to 0.79 while maintaining a similar recall to that of Goel et al. at 0.58, and a slightly increased FNR of 0.42. All these methods rely solely on the textual part of news.

Our proposed model, which utilizes both content (the text part of the news) and comments, represents the culmination of this evolutionary process with a superior accuracy of 0.86, indicating its robustness in correctly identifying fake news. Moreover, it significantly outperforms its predecessors in recall with a score of 0.78, suggesting its effectiveness in retrieving a higher proportion of actual fake news instances. Notably, it has the lowest FNR of 0.19, underscoring its reduced tendency to miss fake news instances.

The precision of our proposed method is slightly lower than the literature best result, but it is still good at 0.93. The F1-score is also strong, showing that our method works well overall. Overall, our proposed method outperforms the existing literature, although there is still room for improvement. Future work could focus on expanding the datasets by incorporating platforms like TikTok and YouTube to further validate the system. Most importantly, we need to find ways to improve recall by analyzing the detail errors while keeping or increasing accuracy

Thus, this model sets a new benchmark for future research and application, highlighting the effectiveness of the evidence fused method in the domain of cross-domain fake news detection.

7. Conclusions

In conclusion, our innovative method for detecting fake news through the integration of content and comments across multiple social media platforms represents a significant advancement in the field. By employing the Dempster–Shafer combination rule, we are able to aggregate and analyze user interactions from diverse sources, providing a more comprehensive view of the information landscape. This approach not only enhances the cross-domain accuracy of fake news detection by 7% compared with existing methods but also offers deeper insights into the dynamics of how news is shared and discussed online. Our findings underscore the importance of considering multiple sources and forms of data in combating fake news, suggesting that such comprehensive methods are crucial for effectively addressing the challenges posed by misinformation in today's digital age. Beyond the dataset limitations, another main limitation of this study lies in selecting all comments for analysis without fully exploring the relationship between parent and child comments. Child comments, which are responses to parent comments or other child comments within the same thread, add significant depth to the conversation. By analysing the interaction between parent and child comments, it becomes possible to identify the most influential parent comments. These comments are crucial as they help guide the direction of discussions and influence the spread of information. The influence of parent comments can be determined by factors such as the number of replies they receive, the sentiment expressed in those replies, and the extent to which ideas originating from the parent comment are shared within the thread. Future research should focus on collecting more datasets from the other social platforms and analyzing how influential parent comments drive discussions and contribute to the dissemination of misinformation, ultimately leading to more effective strategies for combating fake news.

Author Contributions: J.F. contributed substantially to the theoretical formulation, development of the design methodology, dataset creation, experimental design and implementation, analysis, and interpretation of results, as well as the drafting and revision of the manuscript. The remaining authors (J.K., G.K., I.G. and R.D.) provided oversight of the project, contributed to the theoretical framework, assisted with result interpretation, and critically reviewed and revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by 'Research Excellence Scholarship' of Federation University, Australia.

Data Availability Statement: Data will be available with the reasonable request.

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

References

1. Mayfield, A. What Is Social Media. 2009. Available online: https://online.anyflip.com/fmli/xqvr/mobile/index.html#p=8 (accessed on 23 July 2024).
2. Van der Meer, T.G.; Verhoeven, P. Public framing organizational crisis situations: Social media versus news media. *Public Relations Rev.* **2013**, *39*, 229–231. [CrossRef]
3. Boczkowski, P.J.; Mitchelstein, E.; Matassi, M. "News comes across when I'm in a moment of leisure": Understanding the practices of incidental news consumption on social media. *New Media Soc.* **2018**, *20*, 3523–3539. [CrossRef]
4. De Corniere, A.; Sarvary, M. Social media and news: Content bundling and news quality. *Manag. Sci.* **2023**, *69*, 162–178. [CrossRef]
5. Iida, T.; Song, J.; Estrada, J.L.; Takahashi, Y. Fake news and its electoral consequences: A survey experiment on Mexico. *AI Soc.* **2024**, *39*, 1065–1078. [CrossRef]
6. McKay, S.; Tenove, C. Disinformation as a threat to deliberative democracy. *Political Res. Q.* **2021**, *74*, 703–717. [CrossRef]
7. Nguyen, D.; Hekman, E. The news framing of artificial intelligence: A critical exploration of how media discourses make sense of automation. *AI Soc.* **2024**, *39*, 437–451. [CrossRef]
8. Kogan, S.; Moskowitz, T.J.; Niessner, M. Fake news: Evidence from financial markets. *SSRN Electron. J.* **2019**, 3237763.. [CrossRef]
9. Kogan, S.; Moskowitz, T.J.; Niessner, M. *Fake News in Financial Markets*; Social Science Research Network (SSRN): Rochester, NY, USA, 2020
10. USC Scientists Discover the Real Reason Why Fake News Spreads on Social Media—scitechdaily.com. Available online: https://scitechdaily.com/usc-scientists-discover-the-real-reason-why-fake-news-spreads-on-social-media/ (accessed on 22 July 2023).
11. Janicka, M.; Pszona, M.; Wawer, A. Cross-domain failures of fake news detection. *Comput. Sist.* **2019**, *23*, 1089–1097. [CrossRef]
12. Haenlein, M.; Anadol, E.; Farnsworth, T.; Hugo, H.; Hunichen, J.; Welte, D. Navigating the New Era of Influencer Marketing: How to be Successful on Instagram, TikTok, & Co. *Calif. Manag. Rev.* **2020**, *63*, 5–25.
13. Ancu, M. Older adults on Facebook: A survey examination of motives and use of social networking by people 50 and older. *Fla. Commun. J.* **2012**, *40*, 1–12.
14. Parmelee, J.H.; Bichard, S.L. Politics and the Twitter Revolution: How Tweets Influence the Relationship Between Political Leaders and the Public. *Political Sci. Q.* **2013**, *128*, 178–180. Available online: http://www.jstor.org/stable/23563384 (accessed on 30 August 2023).
15. Nguyen, M. Twitter's Role In Politics—northwesternbusinessreview.org. Available online: https://northwesternbusinessreview.org/twitters-role-in-politics-b3ed620465c9 (accessed on 30 August 2023).
16. Utz, S.; Breuer, J. The relationship between networking, LinkedIn use, and retrieving informational benefits. *Cyberpsychol. Behav. Soc. Netw.* **2019**, *22*, 180–185. [CrossRef] [PubMed]
17. 8 Facts About Americans and Twitter as It Rebrands to X—pewrsr.ch. Available online: https://pewrsr.ch/44HbxcN (accessed on 30 August 2023).
18. Takhteyev, Y.; Gruzd, A.; Wellman, B. Geography of Twitter networks. *Soc. Netw.* **2012**, *34*, 73–81. [CrossRef]
19. Donchenko, D.; Ovchar, N.; Sadovnikova, N.; Parygin, D.; Shabalina, O.; Ather, D. Analysis of comments of users of social networks to assess the level of social tension. *Procedia Comput. Sci.* **2017**, *119*, 359–367. [CrossRef]
20. Li, L.; Wen, H.; Zhang, Q. Characterizing the role of Weibo and WeChat in sharing original information in a crisis. *J. Contingencies Crisis Manag.* **2023**, *31*, 236–248. [CrossRef]
21. Japan Social Media Statistics 2023 | Most Popular Social Media Platforms—theglobalstatistics.com. Available online: https://www.theglobalstatistics.com/japan-social-media-statistics/?expand_article=1 (accessed on 30 August 2023).
22. Pérez-Rosas, V.; Kleinberg, B.; Lefevre, A.; Mihalcea, R. Automatic detection of fake news. *arXiv* **2017**, arXiv:1708.07104.
23. Gautam, A.; Jerripothula, K.R. Sgg: Spinbot, grammarly and glove based fake news detection. In Proceedings of the 2020 IEEE Sixth International Conference on Multimedia Big Data (bigMM), New Delhi, India, 24–26 September 2020; pp. 174–182.
24. Saikh, T.; De, A.; Ekbal, A.; Bhattacharyya, P. A deep learning approach for automatic detection of fake news. *arXiv* **2020**, arXiv:2005.04938.
25. Goel, P.; Singhal, S.; Aggarwal, S.; Jain, M. Multi domain fake news analysis using transfer learning. In Proceedings of the 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 8–10 April 2021; pp. 1230–1237.
26. Foster, C.L. Truth as social practice in a digital era: Iteration as persuasion. *AI Soc.* **2023**, *38*, 2009–2023. [CrossRef]
27. Hamed, S.K.; Ab Aziz, M.J.; Yaakub, M.R. Fake news detection model on social media by leveraging sentiment analysis of news content and emotion analysis of users' comments. *Sensors* **2023**, *23*, 1748. [CrossRef]
28. Guo, C.; Cao, J.; Zhang, X.; Shu, K.; Yu, M. Exploiting emotions for fake news detection on social media. *arXiv* **2019**, arXiv:1903.01728.
29. Xu, X.; Li, X.; Wang, T.; Jiang, Y. AMPLE: Emotion-Aware Multimodal Fusion Prompt Learning for Fake News Detection. In Proceedings of the International Conference on Multimedia Modeling, Nara, Japan, 8–10 January 2025; pp. 86–100.

30. Alonso, M.A.; Vilares, D.; Gómez-Rodríguez, C.; Vilares, J. Sentiment analysis for fake news detection. *Electronics* **2021**, *10*, 1348. [CrossRef]
31. Yanagi, Y.; Orihara, R.; Sei, Y.; Tahara, Y.; Ohsuga, A. Fake news detection with generated comments for news articles. In Proceedings of the 2020 IEEE 24th International Conference on Intelligent Engineering Systems (INES), Reykjavík, Iceland, 8–10 July 2020; pp. 85–90.
32. Nan, Q.; Sheng, Q.; Cao, J.; Hu, B.; Wang, D.; Li, J. Let Silence Speak: Enhancing Fake News Detection with Generated Comments from Large Language Models. *arXiv* **2024**, arXiv:2405.16631.
33. Palacios Barea, M.; Boeren, D.; Ferreira Goncalves, J. At the intersection of humanity and technology: A technofeminist intersectional critical discourse analysis of gender and race biases in the natural language processing model GPT-3. *AI Soc.* **2023**, 1–19. [CrossRef]
34. Goldstein, S.; Kirk-Giannini, C.D. Language agents reduce the risk of existential catastrophe. *AI Soc.* **2023**, 1–11. [CrossRef]
35. O'Connor, S.; Liu, H. Gender bias perpetuation and mitigation in AI technologies: Challenges and opportunities. *AI Soc.* **2023**, *39*, 2045–2057. [CrossRef]
36. 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]
37. Ahn, Y.C.; Jeong, C.S. Natural language contents evaluation system for detecting fake news using deep learning. In Proceedings of the 2019 16th International Joint Conference on Computer Science and Software Engineering (JCSSE), Chonburi, Thailand, 10–12 July 2019; pp. 289–292.
38. Safaya, A.; Abdullatif, M.; Yuret, D. Kuisail at semeval-2020 task 12: Bert-cnn for offensive speech identification in social media. *arXiv* **2020**, arXiv:2007.13184.
39. Capuano, N.; Fenza, G.; Loia, V.; Nota, F.D. Content-based fake news detection with machine and deep learning: A systematic review. *Neurocomputing* **2023**, *530*, 91–103. [CrossRef]
40. Pan, J.Z.; Pavlova, S.; Li, C.; Li, N.; Li, Y.; Liu, J. Content based fake news detection using knowledge graphs. In Proceedings of the The Semantic Web–ISWC 2018: 17th International Semantic Web Conference, Monterey, CA, USA, 8–12 October 2018; Proceedings, Part I 17, pp. 669–683.
41. 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, Munich, Germany, 2–4 December 2019; pp. 669–673.
42. Szczepański, M.; Pawlicki, M.; Kozik, R.; Choraś, M. New explainability method for BERT-based model in fake news detection. *Sci. Rep.* **2021**, *11*, 23705. [CrossRef]
43. Kula, S.; Choraś, M.; Kozik, R. Application of the bert-based architecture in fake news detection. In Proceedings of the 13th International Conference on Computational Intelligence in Security for Information Systems (CISIS 2020) 12, Burgos, Spain, 16–18 September 2020; pp. 239–249.
44. Kumar, S.; Kumar, G.; Singh, S.R. Text_Minor at CheckThat!-2022: Fake News Article Detection Using RoBERT. In Proceedings of the CLEF (Working Notes), Bologna, Italy, 5–8 September 2022; pp. 554–563.
45. Stewart, J.; Lyubashenko, N.; Stefanek, G. The efficacy of detecting AI-generated fake news using transfer learning. *Issues Inf. Syst.* **2023**, *24*, 164–177.
46. He, C.; Chen, S.; Huang, S.; Zhang, J.; Song, X. Using convolutional neural network with BERT for intent determination. In Proceedings of the 2019 International Conference on Asian Language Processing (IALP), Shanghai, China, 15–17 November 2019; pp. 65–70.
47. Trueman, T.E.; Kumar, A.; Narayanasamy, P.; Vidya, J. Attention-based C-BiLSTM for fake news detection. *Appl. Soft Comput.* **2021**, *110*, 107600. [CrossRef]
48. Fang, Y.; Gao, J.; Huang, C.; Peng, H.; Wu, R. Self multi-head attention-based convolutional neural networks for fake news detection. *PLoS ONE* **2019**, *14*, e0222713. [CrossRef] [PubMed]
49. Arin, K.P.; Mazrekaj, D.; Thum, M. Ability of detecting and willingness to share fake news. *Sci. Rep.* **2023**, *13*, 7298. [CrossRef] [PubMed]
50. Shrestha, A.; Spezzano, F. Characterizing and predicting fake news spreaders in social networks. *Int. J. Data Sci. Anal.* **2022**, *13*, pp. 385–398. [CrossRef]
51. 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 25th International Joint Conference on Artificial Intelligence: IJCAI, New York, NY, USA, 9–15 July 2016; pp. 3818–3824.
52. Zubiaga, A.; Aker, A.; Bontcheva, K.; Liakata, M.; Procter, R. Detection and resolution of rumours in social media: A survey. *Acm Comput. Surv. (Csur.)* **2018**, *51*, 1–36. [CrossRef]
53. Qian, F.; Gong, C.; Sharma, K.; Liu, Y. Neural User Response Generator: Fake News Detection with Collective User Intelligence. In Proceedings of the IJCAI, Stockholm, Sweden, 13–19 July 2018; Volume 18, pp. 3834–3840.

54. Shu, K.; Cui, L.; Wang, S.; Lee, D.; Liu, H. dEFEND: Explainable Fake News Detection. In Proceedings of the 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining Anchorage, AK, USA, 4–8 August 2019; pp. 395–405.
55. Sharma, D.K.; Sharma, S. Comment filtering based explainable fake news detection. In Proceedings of the Second International Conference on Computing, Communications, and Cyber-Security: IC4S, Delhi, India, 3–4 October 2020; pp. 447–458.
56. Ferdush, J.; Kamruzzaman, J.; Karmakar, G.; Gondal, I.; Das, R. Identification of Fake News: A Semantic Driven Technique for Transfer Domain. In Proceedings of the International Conference on Neural Information Processing, Virtual Event, 22–26 November 2022; pp. 564–575.
57. Ziegele, M.; Breiner, T.; Quiring, O. What creates interactivity in online news discussions? An exploratory analysis of discussion factors in user comments on news items. *J. Commun.* **2014**, *64*, 1111–1138. [CrossRef]
58. Ziegele, M.; Weber, M.; Quiring, O.; Breiner, T. The dynamics of online news discussions: Effects of news articles and reader comments on users' involvement, willingness to participate, and the civility of their contributions. *Inf. Commun. Soc.* **2018**, *21*, 1419–1435. [CrossRef]
59. Sairambay, Y.; Kamza, A.; Kap, Y.; Nurumov, B. Monitoring public electoral sentiment through online comments in the news media: A comparative study of the 2019 and 2022 presidential elections in Kazakhstan. *Media Asia* **2024**, *51*, 33–61. [CrossRef]
60. Raza, S.; Reji, D.J.; Ding, C. Dbias: Detecting biases and ensuring fairness in news articles. *Int. J. Data Sci. Anal.* **2024**, *17*, 39–59. [CrossRef]
61. Hu, B.; Sheng, Q.; Cao, J.; Shi, Y.; Li, Y.; Wang, D.; Qi, P. Bad actor, good advisor: Exploring the role of large language models in fake news detection. In Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 20–27 February 2024; Volume 38, pp. 22105–22113.
62. Micallef, N.; He, B.; Kumar, S.; Ahamad, M.; Memon, N. The role of the crowd in countering misinformation: A case study of the COVID-19 infodemic. In Proceedings of the 2020 IEEE International Conference on Big Data (Big Data), Atlanta, GA, USA, 10–13 December 2020; pp. 748–757.
63. Tanchak, P.N. The invisible front: Russia, trolls, and the information war against Ukraine. *Revolut. War Contemp. Ukr. Chall. Change* **2017**, *161*, 253.
64. Ferdush, J.; Kamruzzaman, J.; Karmakar, G.; Gondal, I.; Das, R. Detecting Fake News of Evolving Events using Machine Learning: Case of Russia-Ukraine War. In Proceedings of the 35th Australasian Conference on Information Systems, Wellington, New Zealand, 5–8 December 2023. Available online: https://aisel.aisnet.org/acis2023/122 (accessed on 30 August 2023).
65. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [CrossRef]
66. Shafer, G. Dempster's rule of combination. *Int. J. Approx. Reason.* **2016**, *79*, 26–40. [CrossRef]
67. Scikit-Learn Developers. sklearn.linear_model.LinearRegression. 2023. Available online: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html (accessed on 23 July 2024).