International Journal of Multidisciplinary Research and Publications
ISSN (Online): 2581-6187

A Review of Research on Detection of False Information on Social Media

Cheng Xiaoting, Li Dongqin, Li Wanjun, Tang Wenfei, Chen Wenya, Lu Kaiyue
College of Management Science and Engineering, Anhui University of Finance & Economics, Bengbu 233030 China
Email address: lidongqin0101@163.com

***Abstract*—*In the context of the rapid popularization of social media, the spread of false information can easily mislead public perception, leading to incorrect judgments and decisions, and eroding public trust in traditional media and news organizations. This has had significant negative effects on the economy, politics, and society. Therefore, there is widespread attention from various sectors regarding false information detection technology. Research on false information detection in social media holds significant practical importance. First, it involves delving into the definition and characteristics of false information in social media from multiple perspectives, analyzing the current situation and its causes. Second, it explores methods for detecting false information in social media using technologies such as large language models and multimodal discourse analysis. Finally, it systematically reviews detection methods based on text, user behavior, and social network structure, and discusses the challenges and future directions in current research.*

***Keywords*—*False information detection; multimodal key; social context perception; knowledge driven; large language model*

I. FOREWORD

With the rapid development of the Internet and social media, the speed and scope of information dissemination have reached unprecedented heights. As an important platform for modern information dissemination, social media has become one of the primary channels for people to obtain news and information due to its low cost, easy accessibility, and rapid spread. However, these advantages of social media also bring significant negative effects, particularly the widespread dissemination of false information. Research on detecting false information on social media is of great practical significance. This paper aims to review and prospectively address the research on detecting false information on social media using technologies such as large language models and social context awareness, providing a theoretical foundation for the development of this field.

II. PROBLEM DESCRIPTION

A. *Definition and Characteristics*

There are various definitions of false information in different studies, but the core conceptusually revolves around the following aspects

- Contrary to Fact: False information is generally defined a s information that does not conform to the actual situatio n or facts. Online false information refers to "information in the network that does not conform to the facts," and f urther categorizes it into types such as defamatory online false information [1].
- Deliberate Dissemination: Many definitions emphasize t hat the spread of false information is for some purpose, s uch as misleading others, manipulating public opinion or gaining an advantage.
- Misleading: A significant feature of false information is t hat it is misleading, that is, to guide the audience to form a wrong perception through incorrect or incomplete infor mation.
- Social Impact: False information is not limited to the con tent itself, but also receives attention because of its impa ct on society.
- Pattern of Manifestation:
  1. Rumors and gossip: Rumors are usually based on unv erified information or completely fictitious content, which spreads rapidly through social media. This kin d of information may involve celebrity gossip, consp iracy theories, etc., which can easily arouse public at tention and discussion[2].
  2. Fake news: Fake news is a deliberately fabricated ne ws report that is usually disguised as real news conte nt in order to attract public attention or manipulate p ublic opinion. This type of information may contain exaggerated, distorted or even completely fabricated facts with the aim of misleading readers or creating c onfusion.
  3. Other forms: include clickbait, deepfakes, misleading headlines and biased news, low-quality news reporting, taking things out of context, fraudulent content, etc. These messages may be generated by negligence, malice or technical means, further exacerbating the spread of disinformation.

B. *Present Situation*

The current state of the spread of false information on social media is a complex and multidimensional phenomenon, with far-reaching impacts on social trust, public discourse, economic activities, and personal privacy. False information spreads at an extremely fast pace and has a wide reach. Studies show that the speed of false information dissemination often exceeds that of true information, leading to the public being exposed to a large amount of unverified information in a short time, which can easily cause misunderstanding and panic. For example, during the pandemic, false information about virus treatments spread rapidly, leading to a decline in public trust and even delays in vaccination and treatment. Moreover, false information can be further exacerbated by deepfake technology, such as forging videos of celebrities or political figures, misleading voters and causing social unrest[3].

The dangers of false information on social media manifest in various aspects: On the societal level, the spread of false information on social media undermines trust and creates social panic. For instance, during the COVID-19 pandemic, misinformation eroded public confidence in health leaders, affecting vaccination rates and the implementation of preventive measures [4]; On the economic level, false information on social media disrupts market order. For example, misleading investment information can lead investors to make wrong decisions, causing stock prices to fall or markets to crash[5] ;On the political level, false information can influence election outcomes by manipulating voter sentiment. For example, during the 2016 U.S. presidential election, false information was used to mislead voters and impact the electoral process [6].

C. Cause of The Incident

- Economic Interest Drive: In order to achieve economic benefits, users of social media platforms tend to deliberately post false information in order to obtain clicks.
- Ideological and Political Purposes: In the context of the rapid development of social media today, the creation of false information is often aimed at manipulating public opinion to achieve certain political goals and to promote certain ideologies. In addition, false information can also be used to create division and exacerbate social conflicts[7].
- Inadequate Detection Technology and Insufficient Regulation: The lack of effective regulatory mechanisms on social media platforms is one of the key reasons for the proliferation of false information. Due to its anonymity and low cost, the spread of false information is more extensive and rapid. Although some platforms have attempted to reduce the spread of false information through algorithmic adjustments and fact-checking, the results have been limited.

III. Challenges And Opportunities Of Social Media False Information Detection Technology

According to different analysis perspectives, virtual information detection methods can be divided into semantic-based detection, social context information-based detection and knowledge-driven detection.

A. Detection Based on Semantic Information Content

The information in social media consists of multiple aspects such as text, images, and audio. Semantic detection based on information content aims to extract features from social media information, primarily mining the characteristics and semantics of text and image modal data, and mapping them into authenticity label detection information for authenticity assessment. Depending on the different types of detection data, it can be divided into content-based semantic detection and multi-modal semantic feature-based detection.

- Content-based Semantic Detection

The most common information carrier on social media is text. This approach focuses on the text itself, using its semantic characteristics to detect misinformation. Early research utilized methods such as neural networks and model training to mine textual semantics, revealing semantic conflicts and narrative inconsistencies in texts. For example, Ma et al [8]. proposed a rumor detection method based on Recurrent Neural Networks (Recurrent Neural Networks, RNN), constructing information on social media as variable-length time series. By learning potential features, this method captures the evolution of context information for related posts over time, effectively identifying false information.

Due to the difficulty of capturing long-distance semantic dependencies in text content with recurrent neural networks, Chen et al. proposed a convolutional neural network (Convolutional Neural Networks, CNN) that combines attention mechanisms and residual networks in [9]. This framework, based on fine-tuned attention mechanisms, captures long-distance dependencies and uses multi-window convolutional neural networks to extract key components and local features, achieving false information detection. Trueman et al[10] .proposed a bidirectional long-short term memory network model based on attention mechanisms, known as (Attention-based Convolutional Bidirectional Long Short-Term Memory,AC-BiLSTM), using C-BiLSTM (Convolutional Bidirectional Long Short-Term Memory, C-BiLSTM) to capture local, global, and temporal meanings of sentences. The attention mechanism then helps remember long input sequences to enhance the detection capability for false information. Additionally, to effectively model the interaction between text sentences, graph neural networks can be utilized to explore fine-grained semantics of texts. For example, Vaibhav et al [11].proposed a model based on graph neural networks in to capture the interaction of sentence information in long news articles.

- Detection Based on Multimodal Semantic Features

The information in social media is composed of text, images, audio, and other elements. Since the complete narrative of an event is primarily made up of text and images, this method mainly extracts and integrates the semantics of text and images to study the consistency characteristics of cross-modal information, thereby detecting false information. The general process of detection is shown in Figure 1.

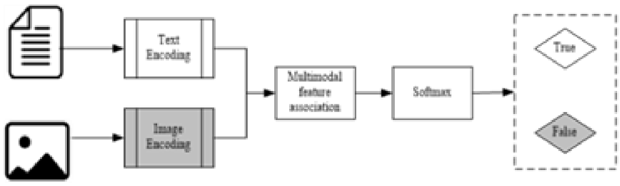

Fig. 1. The basic process of cross-modal semantic association of disinformation detection

By encoding text and images to capture their semantic characteristics, then associating and integrating them, a hybrid model is used for information detection. By revealing the contradictions in semantics across multimodal domains, false information can be identified. For example, Zhou et al[12]. proposed a similarity-aware method for detecting false information (Similarity-Aware Multi-modal Fake News

Detection, SAFE), which uses neural networks to extract and construct news representations from text and visual features respectively. It then delves into the relationships between multimodal features, ultimately jointly learning representations of news texts, visual features, and their interrelations, and predicting the authenticity of news based on these representations. Xue et al. [13] presented a Multi-Modal Consistency Neural Network (MCNN), which is made up of five sub-networks: a text feature extraction module, a visual semantic feature extraction module, a visual tampering feature extraction module, a similarity measurement module, and a multimodal fusion module. This network takes into account the consistency of multimodal data and captures the general features of social media content.

Multimodal feature information has its own focus and complements each other, with complex interactions between information features. By making the most of their complementary nature, they can improve the exchange of information between modalities. Initially, Jin et al[14]. proposed a novel recursive neural network with an attention mechanism (Recurrent Neural Network with an Attention Mechanism, att-RNN) to integrate multimodal features for effective rumor detection. Li et al[15]. introduced an entity -oriented multimodal alignment and fusion network (Entity-Oriented Multi-Modal Alignment and Fusion Network for Fake News Detection, EMAF), which employs entity-centered cross-modal interaction to preserve semantic integrity while capturing details of multimodal entities, thereby detecting fake news through alignment and fusion of multimodal entities. To extract higher-order supplementary information from multimodal context, Qian et al[16]. proposed a hierarchical multimodal context-based attention network (Hierarchical Multi-modal Contextual Attention Network, HMCAN) for rumor detection, modeling multimodal context information and hierarchical semantic information of text into a unified deep model.

Table 1 compares representative detection methods based on multimodal semantic features and their performance on the Weibo dataset. It can be observed that, except for SAFE, all other methods have evaluation metrics exceeding 0.8, indicating that cross-modal information mutual enhancement improves the performance of false information detection; EMAF performs the best, thanks to its entity-centered cross-modal interaction, which captures details of multi-modal entities, thereby enhancing the accuracy of cross-modal association representation.

B. Detection Based on Social Context Information

Social media is composed of diverse entities and their complex relationships, making it difficult to effectively detect false information based solely on semantic features. Social context information refers to the characteristics generated by the dissemination and interaction of information in the network, encompassing elements such as communication patterns and user behavior, which play a crucial role in understanding and evaluating information. Integrating social context information from different perspectives can be divided into detection based on communication structure and detection based on user behavior characteristics.

TABLE 1. Classification and performance comparison of fake news detection methods based on multimodal semantic association

algoritm	Textual features	Image features	Fusion methods	Evaluation indicators	data set
SAFE	Text-CNN	Img2sentence -CNN	Splicing+ Muti-loss	0.790	Weibo
MCNN	BERT	ResNet50	Attention mechanisms+ Muti-loss	0.846	Weibo
att-RNN	BiLSTM	VGG19	Neuron-level attention	0.808	Weibo
EMAF	BERT	Faster-RCNN	Capsule Network	0.893	Weibo
HMCAN	BERT+ Hierarchical coding network	ResNet50	Co-attention networks	0.885	Weibo

- Detection Based on Communication Structure

In social networks, the propagation rules and ways of false information are different from those of real information, so many researchers are committed to combining the propagation structure characteristics and semantic characteristics of information for false information detection.

Early researchers viewed the structure of communication as a text sequence arranged chronologically, processing and learning the time nodes in the communication structure to detect virtual information (Figure 2 provides an example of modeling the communication structure as a text sequence sorted by time). Liu et al[17]. modeled the communication structure as a multivariate time series, constructing a classifier for time series that combines recurrent and convolutional networks to capture features of the communication structure, thereby detecting false information.

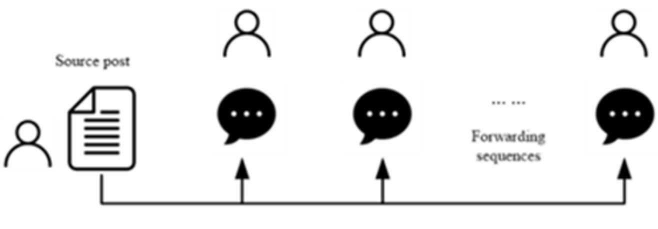

Fig. 2. A time-ordered text sequence modeling approach

Modeling the propagation structure as a tree and using techniques such as recursive neural networks or graph neural networks to extract features of the propagation structure is the current mainstream approach (Figure 3 provides an example of modeling the propagation structure as a tree). Ma et al [18].used a tree-structured recursive neural network, a bottom -up and top-down tree model based on recursive neural networks, for rumor detection. A bidirectional graph convolutional network (Bi-Directional Graph Convolutional Networks, Bi-GCN) was put forth by Bian et al.[19]. This network investigates how rumors spread from both the top-down (top-down) and bottom-up (bottom-up) directions. In the

Cheng Xiaoting, Li Dongqin, Li Wanjun, Tang Wenfei, Chen Wenya, and Lu Kaiyue, "A Review of Research on Detection of False Information on Social Media," *International Journal of Multidisciplinary Research and Publications (IJMRAP)*, Volume 7, Issue 12, pp. 122-128, 2025.

face of increasingly complex false information forgery techniques, integrating multi-level social context information can enable more effective detection. Min et al [20].constructed a heterogeneous propagation tree, which incorporates user information, achieving the integration of various social context information in the field of rumor detection.

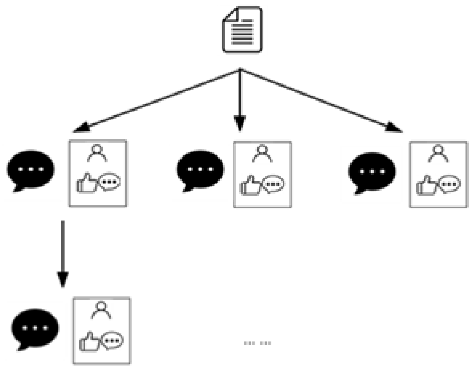

Fig. 3. Modeling method in a tree-like structure

Researchers have a natural advantage in modeling due to the complex interactions of graphs, and have begun to shift towards using graph structures for modeling propagation structures (Figure 4 provides an example of modeling propagation structures using graph-based methods). By changing response characteristics and event structures to data-enhance the propagation graph, extracting meaningful rumor propagation patterns, and learning intrinsic representations of user engagement, He et al. [21] skillfully combined three improvement methods. Sun et al.[22] proposed a novel dual dynamic graph convolutional network (Dual Dynamic Graph Convolutional Networks, DDGCN), which can dynamically model message propagation to better detect rumors.

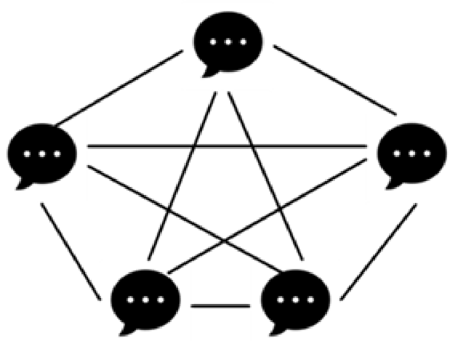

Fig. 4. Modeling method in a graphical structure

- Detection Based on User Behavior Characteristics

The global behavioral characteristics of users in different information are closely related to the authenticity of the information, and integrating user behavior features plays a crucial role in detecting false information. User comments can serve as valuable detection evidence. Shu et al[23]. developed a sentence-comment co-attention sub-network, using an attention mechanism to jointly capture sentences and user comments worth checking, for fake news detection. Wu et al[24]. proposed a decision tree-based co-attention model to select highly credible comments as evidence in a transparent and interpretable manner, thereby detecting false information. Lu et al[25] .developed a graph-aware co-attention network that identifies the authenticity of source tweets by highlighting evidence from suspicious forwards and the words they follow. Literature [26] highlights individual cognitive differences among users, suggesting that the captured evidence is essentially from an individual cognitive perspective and may not accurately reflect objective facts. Therefore, when using user behavior features as a basis for judgment, it is essential to extract high-quality comments, filter out noisy data, and ensure the reliability of the detection results.

*C. K*nowledge-driven detection

Knowledge-driven detection methods require the introduction of relevant background knowledge based on the content of information, comparing known factual information with the information to provide objective evidence for the logical accuracy and authenticity of the information, thereby identifying false information. Depending on the source of knowledge acquisition, these methods can be categorized into knowledge-based inference detection and knowledge-based verification detection.

- Detection Based on Knowledge Reasoning

This method, through pre-training the model, forms rich knowledge within the model as internal knowledge parameters, demonstrating its powerful language understanding and contextual knowledge learning. By exploring these internal knowledge parameters, it extracts the required knowledge facts. Pre-trained language models represented by Transformer's bidirectional encoder (Bidirectional Encoder Representation from Transformer, BERT) [27], leveraging self-attention mechanisms, are widely used for detecting false information (the means of accessing internal knowledge are shown in Figure 5). Shi et al. [28] take a path-based discriminative approach to the issue, utilizing mined rules to assess the veracity of claims, and see it as a link prediction challenge in knowledge graphs. Inspired by the promising performance of pre-trained language models[29], uses two BERT models: one for retrieving potential evidence sentences to support or refute claims, and another for verifying claims based on the predicted set of evidence.

Fig. 5. An implicit knowledge acquisition method based on BERT class pre-trained language models

Represented by Transformer Decoder, large-scale pre-trained models of the ChatGPT class integrate prompt information into the information to be detected through efficient prompt and inference strategy design. They conduct in-depth analysis of the deep knowledge accumulated in their pre-training data (the means of accessing internal knowledge are shown in Figure 6). Jiang et al. 's [30]designs prompts for detecting false information by focusing on elements closely related to factual accuracy in fake news. By invoking parameters from large language models, they submit static element information as prompts to the large model, guiding the large language model to perform logical reasoning, effectively enhancing GPT-4's performance in false information detection.

Fig. 6. An implicit knowledge acquisition method based on chatGPT-like pre-trained language models

Large models struggle to correctly select and integrate the multi-perspective analytical information they provide, which limits their effectiveness as direct detection tools. They can be used as "advisors" for decision-making (as shown in Figure 7). Hu et al [31]. found through experiments that large language models typically can expose fake news and offer ideal multi-perspective theories, but still fall short of fine-tuned BERT. By providing guiding principles from multiple perspectives, they can serve as good advisors to small language models. Literature [32] summarizes expert judgment information as a marker of credibility, using large language models to analyze each credibility marker, and then small language models integrate all the credibility analysis information.

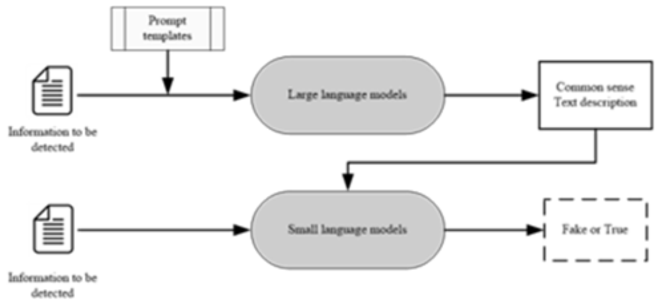

Fig. 7. The big model serves as a research framework for consultants

- Detection Based on Knowledge Verification

This method uses entity linking technology to identify the entity words in the content, and uses external knowledge base (such as knowledge graph, Wikipedia, etc.) to obtain the knowledge closely related to the entity in the text, reveal the potential inconsistencies and false clues. The specific process is shown in Figure 8.

Fig. 8. Entity knowledge acquisition based on an external high-quality knowledge base

Some research designs specifically compare network architectures, using graph neural networks to extract expert information from external knowledge bases, and verify knowledge by comparing the semantic meanings of internal entities with those of knowledge base entities. Hu et al[33]. proposed a novel end-to-end graph neural model that, through a carefully designed entity comparison network, compares the representation forms of contextual entities with their corresponding knowledge base-based representations, to capture consistency between news content and knowledge bases.

Some researchers use knowledge-guided attention allocation to combine the knowledge acquired from external knowledge bases with original information for joint attention processing, focusing on potentially suspicious segments in the source information. Dun et al. [34] proposed a new knowledge -aware attention network that identifies entities in news content and aligns them with entities in the knowledge graph. Entities and their context are used as external knowledge to provide supplementary information, thereby detecting fake news. Chen et al[35] . utilized an attention mechanism to weight article content, proposing an evidence-aware false information detection method. By searching for web articles related to the news text, they ultimately aggregate all features concerning the source of information, the content of web articles, attention weights, and the credibility of web articles to detect false information.

For the task of detecting false information that requires factual verification, a retrieval-enhanced method driven by large models can also be adopted. This approach dynamically searches and integrates supplementary information from outside the large model to provide necessary context for verification. Wang et al[36]. leverage the context learning capabilities of large language models to convert information into first-order logical clauses composed of predicates. Using these large language models, they transform these clauses into specific fact-checking questions, call external tools to retrieve more precise knowledge information, generate corresponding answers, integrate the answers from fact-checking, and provide explanatory information.

IV. Challenges and Opportunities for Social Media False Information Detection Technology

A. *Challenges Ahead*

Although social media fake information detection technology has developed rapidly in recent years, and has been put into use and achieved certain results, it still faces many challenges.

- The Makers of False Information Constantly Adjust Their Strategies to Make It Difficult for Existing Technologies to Detect Them

Nowadays, with the iterative upgrade of false information detection technology, creators of false information are constantly adjusting their strategies to make the forms of false information more diverse and complex in order to evade detection. False information creators use AI technology to generate fake news, fabricate celebrity statements, and synthesize inflammatory speeches by political figures, with quality approaching human levels. This demands that false information detection technology have ultra-high capabilities in distinguishing text fluency and image-video authenticity. In response to individual creators using AI to generate massive amounts of different versions of false content, detection technology must be able to extract core information from large volumes of data and identify underlying logical flaws. Therefore, how to enable false information detection technology to analyze text, images, and videos simultaneously, counter the resistance of false information creators, and detect large volumes of homogeneous content remains a challenge in the field of false information detection.

- The Existing Technology Is Difficult to Detect the Authenticity of Current Hot Events in Time

First, the spread of false information on social media far exceeds the response time of human or AI detection. For example, in 2023, an AI-generated image of the & quot; Pentagon explosion & quot; caused a flash crash in U.S. stocks, which was debunked within 30 minutes but had already caused an impact. This requires detection technology to complete analysis in seconds, yet high-precision models typically require longer computation times. Second, when current hot topics occur, official sources may not release authoritative information, or certain & quot; authoritative accounts & quot; might be maliciously logged into or fabricated, leaving the detection system without a & quot; truth anchor. & quot; False information creators can quickly generate matching false content using AI based on hot topics, while detection models struggle to handle these instantly generated unknown variants. This demands that detection technology have the ability to predict and respond promptly to current hot topics. Therefore, combining text, images, and videos for cross-verification, as well as building a real-time updated knowledge base of hot events, is particularly crucial.

- Over-Testing Can Violate User Privacy and Raise Ethical Issues

User posts containing exaggerated effects, satire, and fringe views on software may be deemed false information by the system, leading users to fear misjudgment and refrain from posting, which in turn suppresses public discourse. This calls for technical personnel to publicly disclose the logic of detection models, allow user appeals, and have multiple authoritative institutions jointly establish detection standards to prevent any single entity from dominating. The detection system needs to analyze users' friend relationships and login devices to track information, potentially infringing on user privacy. Therefore, it is crucial for detection technology to incorporate noise in data analysis to avoid tracking individuals. Thus, finding a balance between verifying the authenticity of information and protecting personal rights is essential.

B. Development Opportunity

- Driven by Policy and Market Demand

In the face of the proliferation of false information caused by AI technology, countries are accelerating legislation to mandate the deployment of AI detection tools on platforms. China's "Regulations on the Governance of Network Information Content Ecosystem" explicitly requires the labeling of AI-generated content, while the EU's "Artificial Intelligence Act" includes deepfake detection as a compliance requirement. These policies not only set technical standards but also foster a vast compliance market. At the same time, the surge in demand from high-value industries is further driving the upgrade of detection technologies. —— Financial institutions are willing to pay premiums for verifying the authenticity of financial reports; Goldman Sachs uses an AI audit system that can identify anomalies in micro-financial characteristics. The healthcare sector relies on detection tools to prevent forged prescriptions and altered images. The dual stimulus from policy and market forces is propelling the development of false information detection technology.

- The Destructive Effect of False Information is Counterproductive to the Development of Detection Technology

The widespread dissemination of false information has become a multi-dimensional social threat, severely impacting social order. Posting false information with racial discrimination or religious opposition under sensitive topics can escalate conflicts, necessitating the use of detection technology to block the release and spread of such information, maintaining social stability. The spread of false information endangers public health; for instance, false disaster information (such as famine, drought, or epidemics) can trigger irrational panic buying and fleeing among citizens, leading to a waste of social resources. It is essential to have false information detection technology to promptly interrupt the spread of false information, quell social unrest, and reassure the public. To foster a healthy social order and ensure that the truth holds the public discourse, we need false information detection technology to combat lies.

V. Conclusion

The identification of misinformation on social media is covered in length in this article. It highlights the severe reality of the spread of false information on social media and further underscores the importance of detection technology for such information. The paper also systematically reviews existing false information detection technologies from three aspects: content semantics, contextual information, and knowledge-driven approaches. However, current detection technologies still face limitations and challenges. In future development, there is potential to research and achieve high-quality, efficient, and highly covert false information detection techniques, fully seizing development opportunities.

Acknowledgment

This work was supported by the Undergraduate Research and Innovation Fund Project of Anhui University of Finance and Economics under Grant No. XSKY25151.

References

[1] Chen Wei, Huo Jungge. Judicial Determination of Maliciously Forwarding False Information on the Internet [J]. Journal of Chongqing University (Social Sciences Edition), 2017,23(5):103-113. DOI: 10.11835/j.issn.1008-5831.2017.05.012.

[2] Wang Xue Ning, Dong Qing Xing, Song Zhi Jun, et al. A Review of Research on the Prevention and Control of Misinformation on Social Media [J]. Journal of Information Resource Management, 2022,12(5):102-113.

[3] Zhao Lu.AI fraud is rampant, how should we fight it? [N]. China Natural Resources News, 2025-4-19 (01)

[4] Navigating the cognitive challenges of distinguishing fact from fiction in today's information landscape. Legroux et al.[2025]

[5] Mastering Avoiding Harmful Examples: Impact on Economy[2023-12-23]

[6] Wang Li. Research and Prospect of Online False Information Detection Technology [J]. Journal of Taiyuan University of Technology, 2022,53(3):397-404.

[7] How and Why Does False Information Spread Online?. Strategian Science.[2020-11-22]

[8] Ma J , Gao W , Mitra P ,et al.Detecting rumors from microblogs with recurrent neural networks[C]//International**Error! Bookmark not defined.Error! Bookmark not defined.Error! Bookmark not defined.** Joint Conference on Artificial Intelligence.2016.

[9] Chen Y , Sui J , Hu L ,et al.Attention-Residual Network with CNN for Rumor Detection[J].ACM, 2019.DOI:10.1145/3357384.3357950.

[10] Trueman T E , Ashok K J , Narayanasamy P ,et al.Attention-based C-BiLSTM for fake news detection[J].Applied Soft Computing, 2021(7553):107600.DOI:10.1016/j.asoc.2021.107600.

[11] Vaibhav V , Mandyam R , Hovy E .Do Sentence Interactions Matter? Leveraging Sentence Level Representations for Fake News Classification[C]//Thirteenth Workshop on Graph-based Methods for Natural Language Processing.2019.DOI:10.18653/v1/D19-5316.

[12] Zhou X , Wu J , Zafarani R .SAFE: Similarity-Aware Multi-Modal Fake News Detection[J].arXiv, 2020.DOI:10.1007/978-3-030-47436-2_27.

[13] Xue J , Wang Y , Tian Y ,et al.Detecting fake news by exploring the consistency of multimodal data[J].Information Processing & Management, 2021, 58(5):102610.DOI:10.1016/j.ipm.2021.102610.

[14] Jin Z , Cao J , Guo H ,et al.Multimodal Fusion with Recurrent Neural Networks for Rumor Detection on Microblogs[C]//the 2017 ACM.ACM, 2017.DOI:10.1145/3123266.3123454.

[15] Li P, Sun X, Yu H, et al. Entity-oriented multi-modal alignment and fusion network for fake news detection[J]. IEEE Transactions on Multimedia, 2021, 24: 3455-3468.

[16] Qian S, Wang J, Hu J, et al. Hierarchical multi-modal contextual attention network for fake news detection[C]//Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval. 2021: 153-162.

[17] Liu Y , Wu Y F B .Early Detection of Fake News on Social Media Through Propagation Path Classification with Recurrent and Convolutional Networks[J]. 2018.DOI:10.1609/aaai.v32i1.11268.

[18] Ma J , Gao W , Wong K F .Rumor Detection on Twitter with Tree-structured Recursive Neural Networks[C]//The 56th Annual Meeting of the Association for Computational Linguistics.2018.DOI:10.18653/v1/P18-1184.

[19] Bian T , Xiao X , Xu T ,et al.Rumor Detection on Social Media with Bi-Directional Graph Convolutional Networks[C]//2020.DOI:10.1609/aaai.v34i01.5393.

[20] Min E , Rong Y , Bian Y ,et al.Divide-and-Conquer: Post-User Interaction Network for Fake News Detection on Social Media[J].Proceedings of the ACM Web Conference 2022, 2022.DOI:10.1145/3485447.3512163.

[21] He Z , Li C , Zhou F ,et al.Rumor Detection on Social Media with Event Augmentations[C]//SIGIR '21: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval.ACM, 2021.DOI:10.1145/3404835.3463001.

[22] Sun M, Zhang X, Zheng J, et al. Ddgcn: Dual dynamic graph convolutional networks for rumor detection on social media[C]//Proceedings of the AAAI conference on artificial intelligence. 2022, 36(4): 4611-4619.

[23] Shu K, Cui L, Wang S, et al. defend: Explainable fake news detection[C]//Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. 2019: 395-405.

[24] Wu L, Rao Y, Zhao Y ,et al.DTCA: Decision Tree-based Co-Attention Networks for Explainable Claim Verification[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics.2020.DOI:10.18653/v1/2020.acl-main.97.

[25] Lu Y J , Li C T .GCAN: Graph-aware Co-Attention Networks for Explainable Fake News Detection on Social Media[J]. 2020.DOI:10.18653/v1/2020.acl-main.48.

[26] Wu L , Rao Y , Lan Y ,et al.Unified Dual-view Cognitive Model for Interpretable Claim Verification[C]//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers).2021.DOI:10.18653/v1/2021.acl-long.5.

[27] Devlin J , Chang M W , Lee K ,et al.BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[J].ArXiv, 2019, abs/1810.04805.DOI:10.18653/v1/N19-1423.

[28] Shi B, Weninger T. Discriminative predicate path mining for fact checking in knowledge graphs[J]. Knowledge-based systems, 2016, 104: 123-133.

[29] Soleimani A , Monz C , Worring M .BERT for Evidence Retrieval and Claim Verification[C]//2020.DOI:10.1007/978-3-030-45442-5_45.

[30] Jiang B, Tan Z, Nirmal A, et al. Disinformation detection: An evolving challenge in the age of llms[C]//Proceedings of the 2024 siam international conference on data mining (sdm). Society for Industrial and Applied Mathematics, 2024: 427-435.

[31] Hu B, Sheng Q, Cao J, et al. Bad actor, good advisor: Exploring the role of large language models in fake news detection[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2024, 38(20): 22105-22113.

[32] Leite J A, Razuvayevskaya O, Bontcheva K, et al. Detecting misinformation with llm-predicted credibility signals and weak supervision. arXiv 2023[J]. arXiv preprint arXiv:2309.07601.

[33] Hu L, Yang T, Zhang L, et al. Compare to the knowledge: Graph neural fake news detection with external knowledge[C]//Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing (volume 1: long papers). 2021: 754-763.

[34] Dun Y, Tu K, Chen C, et al. Kan: Knowledge-aware attention network for fake news detection[C]//Proceedings of the AAAI conference on artificial intelligence. 2021, 35(1): 81-89.

[35] Chen J, Sriram A, Choi E, et al. Generating literal and implied subquestions to fact-check complex claims[J]. arXiv preprint arXiv:2205.06938, 2022.

[36] Wang H , Shu K .Explainable Claim Verification via Knowledge-Grounded Reasoning with Large Language Models[J]. 2023.