

# Chinese AI tool ERNIE Bot Textual Exploration of False Information

Yue Fu\*

Broadcasting and Anchoring School, Communication University of China, Beijing, 100024, China

**Abstract.** This study provides an in-depth discussion of the application of the Chinese AI tool ERNIE Bot in disinformation detection. First and foremost, the spreading characteristics of false information are analysed, especially the phenomenon of rapid spreading and difficulty in distinguishing authenticity on social media. Then, the performance of ERNIE Bot in grammatical, sentiment, and lexical analyses is investigated in detail, revealing the limitations it faces when dealing with complex disinformation. In order to improve the accuracy of detection, this paper proposes countermeasures to improve the AI detection algorithm, enhance data training and model optimisation, and human-machine collaboration. These countermeasures can not only enhance the detection ability of AI, but also give full play to human judgement in information screening, forming a more effective disinformation prevention mechanism. This study provides valuable theoretical support and practical guidance to enhance the application effect of AI in disinformation detection, aiming to escort social information security.

## 1 Introduction

With the rapid development of the Internet, social media and online platforms have become the main channels for the dissemination of information, which, however, also provides a hotbed for the spread of false information. The rapid spread and wide proliferation of false information, especially on social media, has become a global problem. According to CCTV.com, from April to July 2023, the national special operation to combat and rectify online rumours investigated more than 1,600 cases of online rumours and cleaned up more than 420,000 pieces of online rumour information. This data reflects the seriousness of disinformation and its potential threat to society.

The use of Artificial Intelligence (AI) tools is gradually gaining attention in tackling the challenge of disinformation. Baidu Inc. launched ERNIE Bot (Wenxin Yiyuan) in 2023, a tool that not only interacts with users, but also collaborates on creations, and is widely regarded as China's version of ChatGPT. ERNIE Bot has been open for use by global users since 31 August 2023, and its core functions include natural language processing, semantic understanding, and content generation. This makes ERNIE Bot have a wide range of application prospects in the fields of information screening and text analysis.

---

\* Corresponding author: [fuyue@cuc.edu.cn](mailto:fuyue@cuc.edu.cn)

Text analysis, as a key technique in natural language processing, can reveal patterns and structures in text through in-depth analysis of text content. This technique has been widely used in the fields of opinion analysis, sentiment analysis, and topic modelling. As the problem of false information intensifies, how to use advanced text analysis technology to identify and filter false information has become a focus of attention for both academics and the industry. As an advanced AI tool, the potential application of TextCentreOne in disinformation detection is undoubtedly worth exploring in depth.

Disinformation refers to intentionally disseminated false or misleading information that is intended to mislead the public by confusing the public. False information is highly misleading, fast-spreading, and difficult to verify, and often spreads rapidly through social media platforms. Especially during emergencies, political unrest or public crises, the spread of disinformation poses a serious threat to social stability and public safety. The problem of disinformation has become increasingly serious in the country, with the Ministry of Public Security (MPS) investigating more than 1,600 cases of online rumours and cleaning up more than 420,000 pieces of disinformation in a special crackdown from April to July 2023, demonstrating the breadth and complexity of the problem. The proliferation of false information has not only affected public perception, but also impacted the credibility of the government and the media. Therefore, identifying and responding to false information has become an important task in current information management and social governance.

Text analysis is an important technique in the field of natural language processing, which reveals hidden patterns and information in text by analysing the lexical, syntactic and semantic structure of the text. Text analytics technology has a wide range of applications, covering a variety of fields such as public opinion analysis, sentiment analysis, topic modelling, automatic summary generation, and so on. In opinion analysis, text analytics can help enterprises and government agencies to monitor public attitudes and reactions to an event; in sentiment analysis, it can identify emotional tendencies in text and help optimise marketing and customer service. With the increasing size of data and advances in natural language processing technology, the role of text analytics in information processing is becoming more and more important. Especially in dealing with the challenge of false information, text analytics can provide accurate information screening and filtering capabilities through in-depth understanding of semantics, syntax and context, providing technical support for social information security.

## **2 Literature review**

In recent years, the dissemination of fake news and disinformation has become a hot area of academic research, and many scholars have explored the identification and response strategies for this problem from different perspectives. Kumari's study of fake news detection using the BERT model demonstrated the potential of deep learning methods to improve identification accuracy [1]. From multimodal discourse analysis, Benli Wei explored how false information can mislead audiences through multiple mediums and proposed a method for integrating and recognising multimodal information [2]. Nakov et al. developed a series of tasks for identifying fake news in the CLEF-2021 experiments, which emphasised the identification of verification-worthy claims and verified information in different contexts necessity [3].

Shahi et al in a study on the same experimental platform, paid special attention to the refinement of the task of fake news detection, further refining the research methodology in this area [4]. On the other hand, scholars proposed a framework for fake news detection in social media, delved into the application of data mining techniques, and emphasised the challenges of the complex information environments on social platforms [5]. Zhou and Zafarani provided a panoramic view of this research area through a comprehensive survey of

the fake news phenomenon and pointed out the limitations of existing techniques and directions for future research [6].

Mesquita et al. focus on the dissemination of disinformation in the medical field, especially during outbreaks, and discuss how to maintain the truthfulness and accuracy of information in scientific communication [7]. Parikh and Atrey, on the other hand, through a review of multimedia-rich fake news detection, discuss the role of images and videos in the dissemination of fake news, as well as related recognition techniques [8]. Pérez-Rosas et al proposed an automatic fake news detection method based on natural language processing, demonstrating the promising application of machine learning in text analysis [9]. Orso et al provided an in-depth analysis of the phenomenon of fake news proliferation during the COVID-19 outbreak and explored the threat of information flood to public health and its coping strategies [10].

In summary, the research of these scholars provides multi-dimensional perspectives and methods for understanding and responding to fake news and disinformation, covering a wide range of applications from text analysis to multimodal information processing to public health. Together, these studies have advanced the development of fake news identification techniques and provided important theoretical and practical support for academia and society. However, although there have been a large number of studies exploring different methods and application areas of fake news detection, few scholars have specifically investigated the performance of Chinese AI tools, such as ERNIE Bot, in false information detection and its application in text analysis. Therefore, the research selection in this paper fills the gap in this area.

### **3 Methodology**

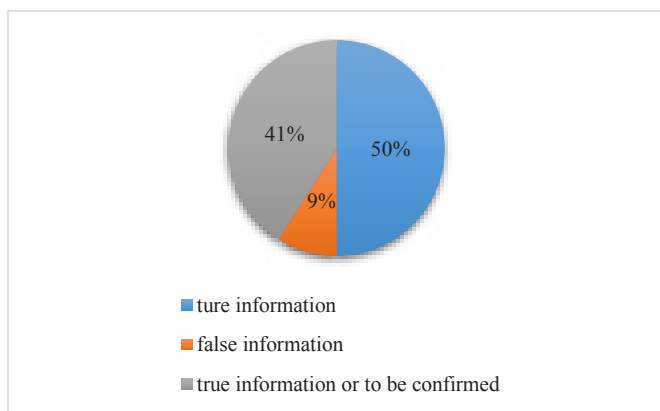
In this study, 100 original pieces of disinformation published by the China Internet Joint Disinformation Platform (<http://www.piyao.org.cn>) between April 2024 and July 2024 were selected for in-depth analysis by means of the case study method. The core of this method lies in the in-depth understanding of the dissemination mechanism and textual characteristics of disinformation through the dissection of specific cases. In the analysis process, firstly, the source, dissemination path, and medium of each piece of disinformation were thoroughly recorded to ensure the accuracy and representativeness of the data. Secondly, the content of these disinformation texts was interpreted one by one to analyse the false statements, misleading information, and emotional incitement contained in them. At the same time, the study also focuses on the characteristics of the dissemination of false information on different social platforms and explores the degree of acceptance and dissemination efficiency of different types of false information among audiences. Through the analysis of these specific cases, the study reveals the common language patterns and structural characteristics of disinformation texts, providing empirical evidence for subsequent prevention and governance.

In terms of text analysis, this study used advanced natural language processing tools - ERNIE Bot and HanLP - to systematically analyse the selected disinformation texts. These tools dig deeper into the linguistic features and structural characteristics of the disinformation through technical means such as segmentation, sentiment analysis and theme recognition of the text. In the specific operation, firstly, the text of the disinformation is sliced and diced by using the segmentation technique to extract the keywords and phrases; secondly, the potential emotional tendency in the text is identified by sentiment analysis to judge whether it is inflammatory or misleading; finally, the main content theme and potential intention of the disinformation are summarised with the help of the theme recognition technique. The results of these analyses not only reveal the linguistic patterns and communication strategies of

disinformation, but also provide technical support for further automated detection and identification, thus effectively enhancing the ability to prevent disinformation.

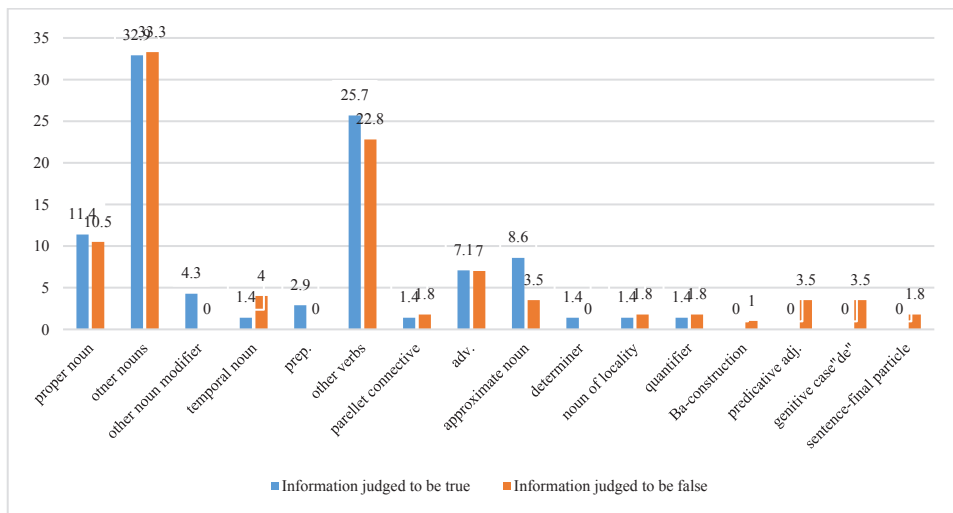
## 4 Results

As shown in Figure 1, ERNIE Bot shows clear classification results in the judgement of 100 pieces of disinformation texts. Specifically, the chart shows that of these 100 texts, 41 were judged to be false information, accounting for 41%. In addition, 9 articles were judged to be true information, accounting for 9% of the total. And more significantly, 50 texts were categorised as true information or expressions that were indeed perceived as true, accounting for 50 per cent of the total. This result suggests that there is some misjudgement in the handling of false information in ERNIE Bot, where half of the texts were incorrectly categorised as true or having true expressions. This misjudgement may affect users' trust when using ERNIE Bot to make judgements about the authenticity of information, so the recognition algorithm for false information needs to be further optimised to reduce the occurrence of misjudgements. The data suggest that there is still much room for improvement in the distinction between false and true information in the current version of the judgement mechanism, and more in-depth research and development is needed, especially in improving the accurate identification of false information.



**Fig. 1.** ERNIE Bot 100 textual judgements of disinformation

As shown in Figure 2, the results of the basic grammar analysis comparing five random texts judged as "true" and "false" are presented. According to the data in the figure, texts judged as "true" show a high frequency in several grammatical structures. For example, subject-verb-object constructions appear 59.41 per cent of the time in "true" texts and 39.68 per cent of the time in "false" texts. Similarly, adjectival constructions are found in 29.77 per cent of "true" texts and 20.93 per cent of "false" texts. In addition, adverbial constructions also show a high frequency of 7% in "true" information, compared to 5.81% in "false" information. This suggests that certain grammatical structures such as subject-verb-object, adjectival and adverbial structures may occur more frequently when determining the truthfulness of a text. In addition, the data in the graph also shows that certain grammatical structures that are less frequently used appear more frequently in "false" texts, such as prepositional structures, which appear 5.81 per cent more frequently in "false" texts than in "true" texts, which appear 1 per cent more frequently. The frequency of prepositional structures is 5.81 per cent in "false" texts and 1.4 per cent in "true" texts. These data suggest that analysing the basic grammatical structure of a text can help to identify the authenticity of the text.



**Fig. 2.** Lexical analysis of five randomly judged 'true' and 'false' information texts

By analysing the qualitative research, this paper observes several significant features of ERNIE Bot in disinformation detection. ERNIE Bot shows high accuracy in processing complex texts. For example, when identifying a piece of false news about "a serious disaster occurred somewhere", ERNIE Bot accurately determines the exaggerated content in the text through semantic analysis and marks it as false information. This ability shows that ERNIE Bot has a strong ability to recognise obviously false information. Second, ERNIE Bot uses a concise and clear language style to help users quickly understand the test results. For example, when confronted with a short false advertising text, ERNIE Bot is able to quickly generate a simple but accurate judgement, such as "This text may contain false ingredients", and suggests that the user should further verify the source. This simple and concise expression makes it easier for the user to understand and adopt the judgement.

Another advantage of ERNIE Bot is that it performs more sensitively in dealing with false information with inflammatory emotions. For example, when analysing a false epidemic text spread by panic, WCS is able to identify the emotional tendency in the text and mark it as possible false information. Last but not least, ERNIE Bot shows some judgement in dealing with false propaganda of limited sales. For example, in a false promotional text about "limited offer", WCS analyses the semantics and context of the text, identifies its false elements, and alerts users to be on the lookout for such marketing tactics. This feature shows that ERNIE Bot also has certain advantages in dealing with complex marketing techniques.

## 5 Discussion

The application of ERNIE Bot in disinformation detection demonstrates significant advantages, but also faces some challenges. The current status of its application shows that although ERNIE Bot can effectively identify and filter some false information, the model is deficient in accuracy and comprehensiveness due to the complexity and continuous evolution of false information. This is mainly due to the uneven maturity of the technology, which results in limited ability to recognise complex disinformation; the bias and insufficiency of data training also affects the detection effect; in addition, the evolution of disinformation propagation strategies makes it difficult for the existing detection methods to adapt to the challenges of new types of disinformation. The application of ERNIE Bot improves the efficiency of information filtering, enhances the public's sense of information security, and

promotes the development of disinformation detection technology. However, the system also faces the problems of misjudgement and omission, which may result in real information being misjudged as false or false information not being effectively identified, while lagging technology updates and data privacy issues also negatively affect its application. To ameliorate these challenges, there is a need to continuously optimise the model to improve the recognition capability by introducing the latest disinformation features and propagation patterns; expand and diversify the training data to reduce the data bias; and at the same time, to ensure the privacy and security of the user's data while improving the detection accuracy, balancing the technology application with the protection of the user's privacy.

In order to enhance the effectiveness of ERNIE Bot in false information detection, the existing system needs to be improved. Firstly, the limitations of technology and the diversity of false information are the main challenges facing the current system. The constant innovation of disinformation techniques makes it difficult for the existing detection models to comprehensively cover all false information; secondly, the insufficiency and imbalance of training data also affect the accuracy of detection; finally, the complexity of dissemination pathways and strategies of disinformation makes it difficult for traditional detection methods to effectively deal with it. To cope with these problems, the following measures are suggested: firstly, optimise the model structure and introduce more advanced natural language processing techniques, such as Transformer and pre-training models, in order to improve the recognition ability of complex disinformation. Second, increase and diversify the training data, reduce data bias by introducing more real and false information samples, and improve the generalisation ability of the model. Finally, strengthen cooperation with social media platforms and information regulators to update the characteristics and dissemination patterns of disinformation in real time to ensure that the detection system can respond to new types of disinformation in a timely manner. Through these improvements, not only can the accuracy of disinformation detection be improved, but also the public's trust in the information and the reliability of the system can be enhanced. However, care also needs to be taken to balance technological advances with the protection of user privacy and to ensure that technological improvements are made while complying with relevant regulations on data protection.

The future application of ERNIE Bot in the field of information screening is promising. The continuous progress of technology and the rapid development of artificial intelligence will promote the expansion of its application in information screening. With the continuous evolution of false information and online fraud, future applications will require more advanced algorithms and models to cope with more complex information environments. ERNIE Bot's deep learning and natural language processing capabilities provide strong support for solving this problem, enabling it to identify more covert and diverse disinformation.

The application of such technology will have a significant positive impact. Improving the accuracy of information screening will help users filter out inaccurate information, improve public information literacy and enhance trust in information sources. In addition, the application of ERNIE Bot will contribute to social stability and reduce the negative impact of false information on public psychology and social opinion. Nonetheless, technological advances may also pose some challenges, such as potential threats to personal privacy and over-reliance on information filtering.

In order to maximise its positive impact while mitigating potential negative effects, measures need to be taken. On the one hand, continue to optimise algorithms and models to improve the accuracy and real-time performance of information screening; on the other hand, pay attention to user privacy protection and ensure that the information processing process complies with ethical norms. Through these measures, ERNIE Bot can play a greater role in information screening in the future and help build a healthier information ecosystem.

## 6 Conclusion

This study investigates the application of Baidu's ERNIE Bot in detecting fake news, focusing on its performance and the challenges it faces. The results and discussion sections reveal both the strengths and limitations of ERNIE Bot in the context of fake news detection. and discussion sections reveal both the strengths and limitations of ERNIE Bot in the context of fake news detection.

The analysis demonstrates that ERNIE Bot performs reasonably well in identifying fake news, with a notable ability to recognise and flag misleading content. The analysis demonstrates that ERNIE Bot performs reasonably well in identifying fake news, with a notable ability to recognise and flag misleading content. Specifically, the tool successfully categorized 41% of the provided fake news samples correctly. However, the study also highlights a significant limitation: a substantial proportion of genuine information was mistakenly classified as fake, accounting for 50% of the samples. discrepancy indicates that while ERNIE Bot is effective in detecting certain types of false information, it struggles with accuracy and precision in This discrepancy indicates that while ERNIE Bot is effective in detecting certain types of false information, it struggles with accuracy and precision in other instances.

The basic syntax analysis further underscores the complexity of fake news detection. The frequent appearance of subject-verb-object structures and the frequent appearance of subject-verb-object structures and specific adjectives in genuine information compared to fake news provides insights into potential linguistic markers that ERNIE Bot could utilize to These findings suggest that while the tool's current capabilities are useful, there is room for refinement in distinguishing between true and false information more reliably. These findings suggest that while the tool's current capabilities are useful, there is room for refinement in distinguishing between true and false information more reliably.

In conclusion, ERNIE Bot shows promise in the realm of fake news detection, leveraging its natural language processing capabilities to identify Nevertheless, the observed limitations, such as the high rate of misclassification and the challenges in handling complex fake news scenarios, highlight the need for ongoing enhancements. Nevertheless, the observed limitations, such as the high rate of misclassification and the challenges in handling complex fake news scenarios, highlight the need for ongoing enhancements. Future improvements should focus on refining detection algorithms and expanding training datasets to bolster the tool's effectiveness and reduce misclassification rates. This approach will enhance ERNIE Bot's ability to address the evolving challenges of fake news and contribute to more reliable information security solutions.

## References

1. S. Kumari, NoFake at CheckThat! 2021: Fake News Detection Using BERT. Computer Science, (2021).
2. B. Wei. Multimodal Discourse Analysis of Misinformation Misinformation. Journal of Guizhou University for Nationalities (Philosophy and Social Science Edition), **152**(4), 193-198 (2015).
3. P. Nakov, G. Da San Martino, T. Elsayed, A. Barrón-Cedeño, R. Míguez, S. Shaar, T. Mandl, The CLEF-2021 CheckThat! lab on detecting check-worthy claims, previously fact-checked claims, and fake news. In Proceedings of the 43rd European Conference on Information Retrieval (pp. 639-649). Springer, (2021).

4. G. K. Shahi, J. M. Struß, T. Mandl, Overview of the CLEF-2021 CheckThat! lab task 3 on fake news detection. In Working Notes of CLEF 2021-Conference and Labs of the Evaluation Forum, (2021).
5. K. Shu, A. Sliva, S. Wang, J. Tang, H. Liu, Fake news detection on social media: a data mining perspective. *ACM SIGKDD Explorations Newsletter*, **19**(1), 22-36 (2017).
6. X. Zhou, R. Zafarani, A survey of fake news. *ACM Computing Surveys (CSUR)*, **53**(5), 1-40 (2020).
7. C. T. Mesquita, A. Oliveira, F. L. Seixas, A. Paes, Infodemia, fake news and medicine: science and the quest for truth. *International Journal of Cardiovascular Sciences*. (2020).
8. S. B. Parikh, P. K. Atrey, Media-rich fake news detection: a survey. in 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR). IEEE. (2018).
9. V. Pérez-Rosas, B. Kleinberg, A. Lefevre, R. Mihalcea, Automatic detection of fake news. *ArXiv*. (2017).
10. D. Orso, N. Federici, R. Copetti, L. Vetrugno, T. Bove, Infodemic and the spread of fake news in the COVID-19 era. *European Journal of Emergency Medicine*. (2020).