

# *The Possibility and Optimization Path of ChatGPT Promoting the Generation and Dissemination of Fake News*

**Li Haoyu**

*Institute of Journalism and Communication, Henan University of Technology, Zhengzhou, 45000,  
China  
935822882@qq.com*

**Keywords:** ChatGPT; AI; Fake news

**Abstract:** With the continuous iteration and gradual maturity of technology, artificial intelligence (AI) is rapidly intervening in various aspects of human production and life, and it has also had a wide and profound impact on news dissemination activities in the media industry. The anthropomorphism and universality of ChatGPT have given people a new understanding of artificial intelligence, but at the same time, it has also brought new possibilities to promote the generation and dissemination of fake news, such as fuzzy model training data, malicious intentions of individual users, and outdated regulatory mechanisms. This requires optimizing the training data of ChatGPT and incorporating manual review, cultivating citizen media literacy and industry professionalism, and continuously promoting the improvement of relevant laws and regulations.

## **1. Introduction**

ChatGPT (Chat Generative Pre trained Transformer) is an artificial intelligence chatbot launched by OpenAI in the United States on November 30, 2022. Essentially, ChatGPT is a Large Language Model (LLM) that utilizes reinforcement learning from human feedback to understand and process human written and natural language, with strong content generation capabilities. ChatGPT can respond to various input questions, engage in multiple rounds of dialogue with users, and assist humans in completing a series of tasks. In other words, ChatGPT's functions are not limited to simple chatting, but can also do homework, write papers, translate, create poetry, and even write code, check program errors, and so on. UBS's research report shows that ChatGPT attracted approximately 123 million active users in less than three months after its launch, making it the fastest growing application in history. Microsoft founder Bill Gates believes that ChatGPT will "change the world".<sup>[1]</sup> However, while technology is changing life, ChatGPT has also brought new hidden concerns about the production and dissemination of fake news.

## 2. Characteristics of ChatGPT Model

### 2.1 Humanistic Simulation: Simulation Model in Interactive Mode

The main source of content generated by ChatGPT is interactive conversations with users, but this does not mean that we can simply classify it as a chatbot. In fact, ChatGPT shows significant differences and progress compared to other traditional artificial intelligence products in terms of the logic, adaptability, and completeness of content generation. The fundamental improvement lies in the Transformer architecture it adopts. Unlike traditional recurrent neural networks (RNNs), Transformer uses a large number of parameters and a large dataset for training. As an advanced deep learning neural network algorithm, it performs well in processing sequence data and implementing self attention mechanisms. Specifically, the advantage of the Transformer architecture lies in its ability to more effectively handle long-distance dependencies, which is crucial for understanding and generating complex natural language texts. This architecture, through large-scale parallel processing, can simultaneously focus on multiple different parts of the text, thus better capturing the richness and diversity of language. Its self-attention mechanism enables the model to pay more attention to important information when processing input data, while ignoring less relevant parts. This not only improves the efficiency of information processing, but also makes the generated content more coherent and accurate. In addition, ChatGPT has formed a reward mechanism by analyzing and sorting the output results of different models, as well as accurately labeling user preferences. This mechanism enables the model to gradually improve its prediction accuracy through continuous iterative training, thereby becoming more closely related to human language habits and cognitive patterns. This process reflects the significant progress of artificial intelligence in simulating human thinking patterns, especially in understanding complex contexts and generating natural language. However, despite ChatGPT's remarkable achievements in simulating human communication, its limitations still need to be recognized. As an artificial intelligence model, ChatGPT does not possess real human emotions and complex thinking abilities. Its response and communication, although seemingly close to real humans, are still products based on algorithms and data. If people overly rely on the virtual field they create, they may still feel lonely and isolated due to a lack of real interpersonal communication and deep emotional connections.

### 2.2 Universality: Intelligent Embedding in Multi Domain

In the field of information retrieval, the deep cooperation between ChatGPT and Bing search engine, as well as the opening up of their interfaces to the outside world, has brought about a technological transformation in the field of information retrieval. This integration is not only a fusion of technology, but also an innovation in artificial intelligence and big data processing methods. Traditional search engines rely on keyword matching and link analysis algorithms to organize, index, and retrieve information. However, this method often overlooks the user's specific query intent and contextual environment, resulting in search results that are not precise and relevant enough. The combination of ChatGPT and Bing breaks this limitation. Specifically, the combination of pre trained large-scale language models (such as ChatGPT) and Bing search engine provides users with a more intelligent and context aware search experience. This experience is not limited to traditional keyword based searches, but can intelligently generate and query relevant and logically coherent content based on the user's specific questioning and conversation history. This means that ChatGPT+Bing is not just searching for existing information, but can create "new" information based on user needs, which is not only more personalized but also better able to meet their actual needs.

In the field of content output, ChatGPT is not limited to a single output mode, but can integrate various domain information and adjust content according to the assigned tasks. For example, in the

field of intelligent customer service, ChatGPT can provide more personalized and accurate answers based on specific customer questions; In the field of game development, it can be used to generate complex game scripts and dialogues; In the field of education and teaching, it can assist teachers in preparing textbooks and designing courses. The later release of ChatGPT-4 represents a significant transformation of artificial intelligence from single modal models to multimodal models, and achieves a leap in content processing capabilities. For example, ChatGPT-4 can analyze image content and generate relevant text descriptions, or generate corresponding images based on text information. This multimodal processing capability greatly expands the application scope of artificial intelligence in various fields, bringing more possibilities to human work and life.

### **2.3 Efficiency: Instant Connection under Technological Development**

Starting from the underlying logic of technology, exploring the model physique and advantages of ChatGPT, it is not difficult to find that acceleration is a prominent feature of AI technology. On the one hand, the immediacy of AIGC content production accelerates the connection between people and information. ChatGPT can complete various types of texts such as paper writing, news production, poetry creation, code programming, and information gathering within seconds based on human dialogue instructions, improving the timeliness and effectiveness of information acquisition. On the other hand, the rapid development of ChatGPT is also accelerating the rapid evolution of artificial intelligence technology. Since the release of GPT by OpenAI in 2018, there have been 5 rounds of iterative upgrades, and behind each technological update is the feeding of larger parameter content and the accelerated stacking of numerous technologies. For example, the beta version of ChatGPT used 10000 Nvidia GPUs to train models, and the new generation GPT-5 model is currently being trained on 25000 Nvidia GPUs. It can be said that every aspect of ChatGPT's birth to this day cannot do without the strong support of tens of thousands of Nvidia and other computing devices. Undoubtedly, ChatGPT has completed the integration of human knowledge to a greater extent, with high-quality natural language interfaces that accelerate the connection between content, technology and technology, and is a greater connection between things. In the future, AI technology will continue to drive the arrival of the "Internet of Things" era in the accelerating modern social system.

## **3. ChatGPT Promotes the Possibility of Generating and Disseminating Fake News**

### **3.1 Fuzzy Model Training Data Leading to Output Distortion**

The quality issue of the training dataset is a key factor in the distortion of output results. The training of the model mainly relies on a large amount of text data, so these text data are the foundation for the model to learn knowledge and establish predictive ability. However, the quality of text data on the Internet is uneven, including a large number of errors, false and misleading information. These inaccurate information will be learned by the model and added to its knowledge base, thereby affecting the accuracy and reliability of the model in generating news. At the same time, the limitations of the model's generalization ability are also an important factor leading to distorted output results. Although artificial intelligence models such as ChatGPT have achieved significant results in processing natural language tasks, their generalization ability is still limited and there is still a significant gap compared to real humans. This means that when dealing with unseen, domain specific, or novel contextual problems, the model may not be able to provide the correct answer due to the lack of innate information and independent thinking ability. At the same time, the learning model of ChatGPT may also fall into an overfitting state during the training process, which means that irrelevant information is fused together during processing, resulting in chaotic output information.

Meanwhile, intentional or unintentional human factors may also lead to errors in model training. In the data preprocessing stage, engineers need to perform data cleaning, annotation, and filtering to improve data quality. However, due to the subjectivity and errors of human judgment, some inaccurate information may still enter the training set as a result. Moreover, during the model optimization and debugging phase, selecting inappropriate parameters and loss functions may also lead to the model overly focusing on certain features, thereby affecting the accuracy of the output results. <sup>[2]</sup>

### **3.2 Individual Users Use Technology to Maliciously Spread Fake News**

In the current information age, the speed and scope of dissemination far exceed any previous period, and although ChatGPT is an advanced natural language generation model with strong text generation capabilities, there is also a risk that this technology may be maliciously exploited by individual users to create and disseminate fake news. Compared to other tools based on artificial intelligence technology for automated generation, the ChatGPT language generation model has higher quality in terms of syntax, logic, and consistency. This makes the content generated by it more persuasive, thereby increasing the recipient's trust in the information it produces. In addition, the difficulty of distinguishing between intelligently generated text and human written content has increased, making the challenge of identifying authenticity even more severe. At the same time, because of the extremely fast speed and wide coverage of information dissemination on the Internet. Individual users can take advantage of this feature to maliciously induce ChatGPT to generate misleading or fabricated articles, and quickly spread the articles on social media and other platforms, thereby expanding their influence. Compared to the previously artificially fabricated fake news, this carefully crafted false information can provide more effective ideas for public interaction, thereby making fake news more fully connected to the public and participating in their social relationships. <sup>[3]</sup>

### **3.3 Failure of Regulatory Mechanisms Leading to Sustained Dissemination of Fake News**

In the past era of traditional media, traditional news organizations played a role as "gateways" in the process of information screening and dissemination. However, in today's digital media environment, it is difficult for many different sources of information to accept the same level of censorship, which gives false news greater penetration space. From a technical perspective, artificial intelligence models such as ChatGPT have powerful learning and generation capabilities, which means they can efficiently generate various types of content, including fake news. Therefore, although the development of technology has brought convenience to society, relevant regulatory authorities have not yet kept up with the times. In the absence of effective regulatory mechanisms, malicious users are more likely to use these technologies to create and disseminate false information, thereby causing public panic, inciting public emotions, or misleading public opinion.

At the same time, social media has also played a role in promoting the spread of fake news. Due to the diverse and fast updating sources of information on social media, it is difficult for users to distinguish authenticity. In addition, algorithmic recommendation systems often prioritize the personalized needs of users as the primary optimization objective, and this single dissemination mechanism can exacerbate the information cocoon effect. <sup>[4]</sup> In this context, fake news is likely to be disseminated more widely through algorithmic recommendation systems. Therefore, the lack of effective regulatory mechanisms will make news platforms unable to timely identify and filter false information, thereby increasing the risk of the spread of false news.

## 4. ChatGPT Promotes the Governance Path of Generating and Disseminating Fake News

### 4.1 Optimize Training Data and Incorporate Manual Review

The quality of training data is one of the key factors affecting model performance, and ChatGPT's corpus system not only has high-quality annotated text data collected and processed through methods such as crawling from open-source code libraries, expert annotations, and user submissions, but also numerous Open AI channels such as books, magazines, encyclopedias, forums, etc., and a large amount of unlabeled text data formed after preliminary cleaning.<sup>[5]</sup> These data may contain incorrect, outdated, or biased content, leading to the model learning inaccurate or harmful knowledge. To address this issue, we can start from three aspects: data filtering and cleaning, data sources, and data annotation. Firstly, effective data filtering and cleaning strategies can be designed before data collection, in order to eliminate data that does not meet quality standards and ensure the diversity and representativeness of training data, thus avoiding excessive focus on a single field. In terms of data sources, reliable ones should be chosen as much as possible to improve the quality of training data. When collecting training data, it can be obtained from authoritative and reliable channels, such as official published data, reports from academic research institutions, etc. Finally, effective annotation can be performed on the training data to help the model better understand semantic structures and improve the accuracy of generated content.

However, optimizing training data alone cannot completely eliminate the possibility of output distortion. Therefore, introducing manual review as an auxiliary means has also become a necessary measure. For example, before the model is officially launched, experts or relevant professionals can be invited to evaluate and review the content generated by the model to ensure that it meets the expected results. After the artificial intelligence model goes live and runs, a real-time monitoring system can also be designed to capture potential problems. Once non-compliant or harmful content is found, information output can be immediately stopped and corresponding adjustments can be made. At the same time, a user feedback mechanism can be established to encourage users to actively provide feedback on generated content, and to continuously optimize and improve the model based on user feedback.

### 4.2 Improving Public Media Literacy and Promoting Media Self-discipline

Compared with the era of traditional media, the subject of news production and dissemination in the era of artificial intelligence is gradually becoming more generalized, and every individual may become any link in the news production and dissemination chain. Therefore, media literacy should not only be a requirement for practitioners, but also a responsibility for every public living in the present. In the traditional public perception, the media, as the disseminators and channels of news information, has a much higher credibility of information than certain relevant interest organizations or individuals, which allows some unscrupulous individuals to take advantage of it. Therefore, media literacy education should help the public realize that there is no distinction or qualitative distinction between good and bad artificial intelligence technology itself. The key lies in the value pursuit and related operations of various subjects using artificial intelligence technology, as well as the various motivations hidden behind their use of intelligent technology.<sup>[6]</sup> From this, it can be seen that the public cannot fully accept news information, but should learn to examine the content, discover the hidden motives and purposes behind it, and engage in calm and critical thinking.

The potential negative impact of ChatGPT generating fake news cannot be eliminated solely by the public, and it is necessary for the media to enhance their industry self-discipline. Only when the awareness of industry ethics becomes an internalizing force in the hearts of media practitioners to guide behavior, an internalization under the premise of punishment, and a transformation from



mandatory compliance to conscious compliance, can the self-discipline of new media truly be achieved. Therefore, new media workers should consciously maintain the authenticity and objectivity of communication, firmly oppose unethical behavior of falsification, make practitioners consciously and actively uphold social responsibility, and make social responsibility an internalized force of new media. [7]

### 4.3 Improve Relevant Regulations and Increase Punishment Efforts

In the current digital age, the phenomenon of using artificial intelligence such as generative pre training models (such as ChatGPT) to produce and disseminate fake news is constantly increasing. This poses a great challenge to information security and seriously disrupts the order of cyberspace. Therefore, the governance of its chaos is urgent. Firstly, at the regulatory level, corresponding regulatory policies and laws and regulations should be formulated based on the technical and industrial characteristics involved in artificial intelligence such as generative pre training models. By clearly defining the concepts of fake news, false information, and harmful content, and standardizing their identification and judgment standards, we provide a basis for relevant administrative management departments to carry out their functions. The regulations should also emphasize the responsibilities of information publishers, platform operators, developers, and other parties in preventing and combating fake news. For example, requiring publishers to verify the authenticity of information; Require the platform to review and control risks of published content; And require developers to optimize algorithms to reduce the generation of harmful information.

While improving laws and regulations, it is also necessary to strengthen the punishment for using artificial intelligence tools such as ChatGPT to generate and disseminate fake news. Strict crackdowns should be imposed on illegal activities. Punishment can be imposed at multiple levels, including fines, suspension of business, and revocation of licenses, by establishing comprehensive punishment measures and rules. In addition, criminal accountability procedures can be initiated in a timely manner to address the severity and social impact of illegal acts. In the specific implementation process, the administrative management department should also closely cooperate with the judicial organs, improve the understanding and processing ability of law enforcement personnel in relevant fields through professional training and technical support. It should be noted that fairness, openness, and transparency should be ensured in the law enforcement process to ensure that regulatory effectiveness is widely recognized.

## 5. Conclusions

Although the success of ChatGPT has raised expectations for the development of artificial intelligence, we still need to approach them rigorously and pragmatically. Technology itself is a double-edged sword, which not only promotes the progress and development of the communication ecosystem, but also brings more flaws such as fake news output. Therefore, it is necessary to maintain cold thinking in the midst of technological fever, fully stimulate human consciousness and subjectivity, and stimulate creativity. So as to rationally utilize the advantages of digital technology and achieve efficient integration of human and intelligence.

## References

- [1] Zhou Hongyu, Li Yuyang. *The Impact and Response Strategies of ChatGPT on the Education Ecology [J]. Journal of Xin Jiang Normal University (Philosophy and Social Sciences Edition), 2023, 44 (04): 102-112.*
- [2] Lu Jingwei, Guo Chao, Dai Xingyuan, Miao Qinghai, Wang Xingxia, Yang Jing, Wang Feiyue. *Q&A After ChatGPT: Opportunities and Challenges of Ultra Large Pre training Models [J]. Journal of Automation, 2023, 49 (04): 705-717.*

- [3] Kuang Wenbo. "Conspiracy" and Avoidance Paths of Fake News in the Era of Artificial Intelligence [J]. *Journal of Shanghai Normal University (Philosophy and Social Sciences Edition)*, 2019, 48 (04): 104-112.
- [4] Peng Lan. Multiple factors leading to information cocoons and the path of "breaking the cocoon" [J]. *Press Circles*, 2020 (01): 30-38+73.
- [5] Qian Li, Liu Yi, Zhang Zhixiong, etc Technical Foundation Analysis of ChatGPT [J]. *Data Analysis and Knowledge Discovery*, 2023, 7 (03): 6-15.
- [6] Jiang Bowen. Artificial Intelligence and the Governance of Fake News on Social Media [J]. *Media*, 2021, (14): 61-63.
- [7] Shen Yue'e. Lack of Ethics in New Media and Its System Construction [J]. *Gansu Social Sciences*, 2012, No. 197 (02): 19-21+41.