Research and Evaluation Methods for AI Applications in the News Media Industry

Honglei Cheng[1], Peizhi Zhao[2], Xinyu Ding[2], Yuhao Gao[2], Xichen Cao[2], Hui Liu[2]

[1]Tianjin Foreign Studies University, Tianjin, China

[2]China University of Petroleum (East China), Tsingtao, China

Keywords: artificial intelligence; news media industry; AI macromodels; evaluation system; digital consequential

Abstract: With the rapid development of artificial intelligence technology, more and more AI macromodels are put into the market and widely used, which has a great impact on the news media industry. In this paper, we first investigated the current status of the application of AI macromodels in the news media industry and the many challenges it faces, and based on more than 100 AI macromodels built in the OpenRouter platform, we established an evaluation system for the content generation of AI macromodels according to the four aspects of performance, neatness, convenience, and content, and compared the textual content generation capability of the mainstream macromodels in the news media industry, and analyzed that the application of AI macromodels in the news media industry has a bright future, but more details and hints are needed to generate more accurate content, and finally puts forward the outlook of the future digital consequential AI scoring based on the existing analysis results.

1. Current Application Status

At present, news media organizations in China and the West have widely introduced AI to serve the main links of news production, and there are many similar situations in the use scenarios. However, there are also differences in the application in some specific links [1]. In comparison, Western media pay more attention to the information collection capabilities of AI, while mainstream media in China pay more attention to the public opinion monitoring functions of AI [2]. In terms of news writing, Western media have trained a large number of automatic news writing models and have widely put them into the practice of news. In terms of news distribution, artificial intelligence technology has a significant advantage in the level of personalized push, and media at home and abroad have fully utilized the technical superiority of AI at this level. Although AI technology has shown great potential in improving the efficiency of news collection and writing, it still has shortcomings in content depth, personalization, and information authenticity. This article aims to establish a comprehensive evaluation system to systematically evaluate the application of AI large models in the news media industry and explore how to optimize their performance to better serve the news media industry.

2. The Superiority of AI Technology in the News and Media Industry

2.1. Intelligent News Information Collection

News collection, that is, the activity of gathering information during the news interview process, is an extremely important link in the news reporting process, and is the foundation and premise for writing good news reports [3]. There is a saying in the journalism world: "Seven parts interview, three parts writing." In the traditional field of editing, reporters usually rely on personal interviews with others to collect news materials. Although this collection method has unique advantages, its application is limited to a specific range. When the interview task exceeds the scope, some valuable news will be neglected due to the limitations of manual information collection. However, with the application of artificial intelligence technologies such as voice recognition and visual recognition in the news industry, news collection has broken through the traditional manual collection method and

Copyright © (2024) Francis Academic Press, UK DOI: 10.25236/icfmhss.2024.021

entered the stage of intelligent collection[4].

2.2. Intelligent News Writing

When a large number of news leads are obtained, reporters need to organize and filter them before entering the news writing process, which is very laborious. Taking social reporters as an example, sometimes they may cover multiple stories in a day and publish at least one article daily. The length of an article is about 1,000 words, and the word count of a manuscript can vary from 500 to 3,000 words [5]. When major events occur, the number of drafts can even increase several times. Therefore, the emergence of writing robots is in line with the times. Its appearance not only signifies the further optimization and development of artificial intelligence technology but also reflects the actual needs of the current media industry. As early as 2014, the Associated Press collaborated with Automated Insights to launch the news robot Wordsmith, which is used for automated news writing and can generate millions of articles per week [6]. In China, robots such as Tencent's Dreamwriter, Baidu's Writing-bots, First Financial Daily's DT Editor, and Toutiao's Xiaomingbot represent the early development of automatic news writing technology [7].

The core of robot writing is natural language processing, which also involves multiple artificial intelligence technologies such as data mining, machine learning, search technology, and knowledge graphs. When algorithms or computer programs automatically collect information, they automatically process the input or collected data content, and finally compile it into news text for publication [8]. However, current writing robots do not possess the ability to think independently like humans. They can only complete the task of drafting and writing manuscripts based on pre-filled databases and established algorithmic programs. Therefore, they are only suitable for simple information in fields such as sports and finance, and it is difficult to make breakthroughs in the depth and personalization of writing. These shortcomings make the current robot writing still in its infancy [9].

2.3. Intelligent News Broadcasting

At present, "AI synthetic anchors" are applied in the broadcast of news content, offering advantages that traditional announcers and hosts cannot match. On one hand, no matter what the news is, it can broadcast at any time and place, without any complaints or "emotions"; on the other hand, it "does not need rest," which can greatly ensure the timeliness of the news. It can also use "countless clones" to provide news reports at multiple sites simultaneously, not only greatly reducing the workload of announcers and hosts but also enabling news to be broadcast as soon as it arrives, effectively improving the efficiency of news production. However, current AI anchors do not possess "self-awareness" and lack the ability to empathize with people. The news they broadcast only achieves a simple correspondence presentation of text and database audio-visual materials, and they cannot understand and express the hidden meanings of the text, the feelings, attitudes, and corresponding emotional changes caused by news events. Compared with traditional announcers and hosts, "AI synthetic anchors" lack professionalism, flexibility, and adaptability, and there are issues such as mechanical pauses that are difficult to understand.

3. The Three Major Challenges of AI Empowerment in the Production of the News Media Industry

3.1. Mediocrity and Bias in Content

Although artificial intelligence has made significant progress in generating text, the content it produces is often just an objective description of events, lacking in-depth exploration and unique insights, and cannot delve into the background and inside stories of news events like human journalists. In addition, due to the lack of ability to directly contact the scene and relevant personnel, artificial intelligence finds it difficult to obtain real on-site information and details when reporting, leading to more generalized content. Therefore, while artificial intelligence can assist to some extent in news reporting, it still cannot replace the core competitiveness and importance of human journalists in the field of news [10]. On the other hand, artificial intelligence can only write texts based on existing

information on the Internet, which means that the content it outputs must not be first-hand information and original content, making it difficult to ensure the increment of information, and thus the value of the news is greatly reduced.

Furthermore, due to the inherent limitations of artificial intelligence technology, there may be a certain bias when generating text, which can affect the objectivity and fairness of the news [11]. For example, the recent popular artificial intelligence image generation models such as Dall-E and Stable Diffusion have amplified stereotypes, especially in terms of gender discrimination and cultural bias. When prompted to create images of "cleaning staff," these artificial intelligence image generation models generate all images as female, but if asked to generate "engineers," all generated images are male.

3.2. Privacy and Copyright of Data

A substantial amount of personal user information is collected into the corpus for machine learning by large language models such as ChatGPT without the knowledge of the users or platforms, violating privacy data protection laws such as the European General Data Protection Regulation (GDPR). As a result, Italy has banned the use of ChatGPT within its territory, and other European countries like the United Kingdom and Ireland have initiated investigations into ChatGPT's data privacy issues [12].

Furthermore, AI often encounters disputes in terms of data and information copyright. Within less than half a year since its inception, ChatGPT has been embroiled in disputes with multiple content platforms. The technology news website CNET used ChatGPT to write several articles, which were not only pointed out to have multiple errors but were also found to have plagiarized other protected information sources. The social platform "Reddit" also announced that it will charge companies like Microsoft, Google, and OpenAI for data usage fees, as these companies have used Reddit's platform content to train AI chatbots through their application programming interfaces.

3.3. Authenticity and Fabrication of Information

1) Fake Textual Information

The emergence of artificial intelligence models for automatic content generation has made the creation of false information more rapid and deceptive. They can generate a large amount of false information in a short time and widely disseminate it through online channels [13]. Moreover, they can mimic human styles to polish the content, making the produced content more realistic, so that the public finds it difficult to distinguish between truth and falsehood, as well as the source. A lawyer in California, USA, asked the artificial intelligence chatbot ChatGPT to generate a list of legal scholars who had committed sexual harassment. The program output a message about sexual harassment by a professor at George Washington University and cited an article in The Washington Post in March 2018 as the source of the information. However, in fact, the professor had never been accused of sexual harassment, and the cited report from The Washington Post did not exist. When a reporter asked Microsoft's "Bing," supported by GPT-4, the same question, it again fabricated a false report about the professor, citing a column article he published in USA Today.[14] Similarly, ChatGPT recently made up news that an Australian politician was jailed for bribery, and the politician is suing OpenAI for defamation, which will be the world's first lawsuit against an AI robot [15].

2) Image fraud

The proliferation of artificial intelligence image generation technology has led to a rampant phenomenon of image forgery. In April 2023, several images generated by the AIGC platform were widely circulated on global social networks: one showed a scene of former U.S. President Trump being arrested, another showed the Pope wearing a fashionable white down jacket, and another showed Elon Musk holding hands with the CEO of General Motors. These images look almost identical to real news photography, and without prior notification that they were generated by artificial intelligence, it is difficult for the audience to discern their authenticity. The visual false information brought by the artificial intelligence-generated image models has a high degree of concealment and incitement. On the one hand, similar to the previous "deep fakes," these images confuse the audience through visual forgery, and due to their high realism, the public finds it difficult to identify the truth. On the other hand, compared with text forgery, image forgery and other multimodal content forgeries

are more intuitive and impactful, which may have a stronger emotional impact on the audience. Therefore, this kind of visual multimodal forgery may bring a more serious social crisis. What is more noteworthy is that unlike visual forgeries made by Photoshop or deepfake software, artificial intelligence image generation models can quickly generate visual media information within seconds by inputting a few simple text prompts, and spread rapidly through social media platforms [16]. This "digital wildfire" of false information can quickly spread false images or videos through social media platforms, achieving a viral distribution.

4. AI-Empowered News Media Industry Evaluation Analysis - Based on OpenRouter

Through the analysis of the application of AI in the news media industry as mentioned above, a preliminary decision has been made to establish an evaluation system based on four aspects: performance, neatness, convenience, and content, to comprehensively assess the capabilities of AI macro-models in content generation within the news media industry.

4.1. Introduction to OpenRouter

OpenRouter is an intermediary agent that integrates a variety of large models, and it can be directly accessed domestically, as shown in Figure 1. Through OpenRouter, you can call over 100 excellent large models currently available on the market, including the popular OpenAI's ChatGPT series, Anthropic's Claude series, Google's PaLM and Gemini series, etc.[17]

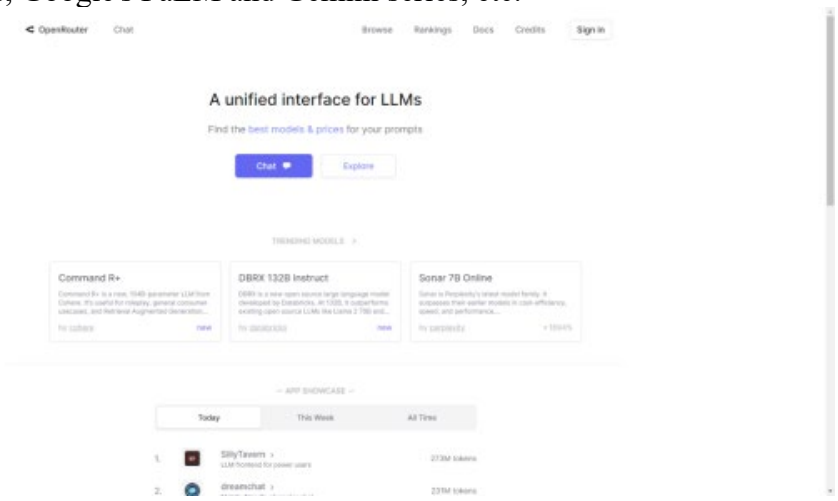

Figure 1 Main Page of the OpenRouter Website

Clicking the "Chat" button on the main page of the website will take you to the chat page where you can "communicate" with the large models. The red-boxed section allows you to select and add over 100 different large models; the yellow-boxed section displays the models that have been added; the blue-boxed section is the chat box where you can input text according to your own needs; the green-boxed section is the dialogue box where the AI large models will return the answers based on the questions asked.

4.2. Performance Analysis

This application example uses the topic "Write a press release on the Chinese table tennis team winning the championship at the Rio Olympics" and invites 8 AI large models including GPT-3.5 Turbo, GPT-4, and Mistral 7B Instruct to generate press releases.

The OpenRouter website has a corresponding "tokens speedometer" for each AI model during the content generation process. For example, the MythoMax 13B model has a generation speed of 19.2 tokens/s, which means it generates 19.2 symbols per second; the Mistral 7B Instruct model has a generation speed of 19.9 tokens/s, meaning it generates 19.9 symbols per second; the GPT-4 model has a generation speed of 6.9 tokens/s, which translates to 6.9 symbols per second. In summary, the performance of different large models can be evaluated based on the symbol generation speed of each model.

4.3. Neatness Analysis

Except for the GPT-4 model, which generates content in the format of: Title-Content, other large models do not generate a title by default and directly generate the content. Subsequently, the neatness of the content for large models can be analyzed based on the output of other document formats, such as letters, holiday greetings, invitations, etc.

4.4. Convenience Analysis - Chinese Language Habits

The four models, MythoMax 13B, Mistral 7B Instruct, GPT-4, and Neural Chat 7B v3.1, received instructions in Chinese and the output content is also in Chinese; whereas the Capybara 7B(free) and Mistral 7B Instruct(free) models received instructions in Chinese but the default output content is in English. Further experiments can be conducted to analyze the convenience of using Chinese language habits with existing large models.

4.5. Content Analysis

1) Coherence Analysis

Both MythoMax 13B and GPT-4 have output coherent Chinese text.

However, Mistral 7B Instruct and Neural Chat 7B v3.1 exhibited logical confusion and incoherent statements.

2) Accuracy Analysis

The content generated by GPT-4 did not include detailed information such as the match schedule, players, and scores. Instead, it focused on the description of "China's team winning," which met the input requirements of writing a press release on "China's table tennis team winning the championship at the Rio Olympics."

In contrast, it shows that the MythoMax 13B model, while satisfying the condition of "China's team winning," fabricated details such as the opponents, participants, and scores in the match.

5. Conclusion

This paper has thoroughly investigated the current state of AI applications in the news media industry and the challenges faced. Based on the large model platform provided by the OpenRouter website, an evaluation system for AI content generation has been established, encompassing four aspects: performance, neatness, convenience, and content. By deeply testing the generation capabilities of various AI large models with the task of "writing a press release on China's table tennis team winning the championship at the Rio Olympics," it was found that: the GPT-4 model, which excels in content generation, has a notably slower generation speed; other AI large models besides GPT-4 all exhibit a lack of neatness in the content generated; a considerable number of AI models do not fully accommodate the convenience of using Chinese, accepting Chinese input but outputting in other languages; a significant portion of AI large models have issues with incoherence and inaccuracy in the content generated. Moreover, to generate more accurate content with AI large models, more details and prompts are needed to prevent the models from "self-extending," fabricating content, or making up information arbitrarily.

Subsequently, an evaluation of AI large models can be conducted using a scoring system that assigns weights to performance, neatness, convenience, and content, providing a more intuitive numerical score for assessment.

References

[1] Tang Yuwei, Liu Yongfeng. A Review of the Development of Artificial Intelligence in the Field of News Communication [J]. Western Radio & Television, 2018 (5).

[2] Ren Ruijuan, Wang Baochao, Zhao Yaxian.Evolution and Trends: The Application of Artificial Intelligence in the Field of Media [J]. Reviews of Journalism and Communication, 2021 (2).

[3] Chan-Olmsted S M. A review of artificial intelligence adoptions in the media industry [J].

International journal on media management, 2019(3-4).

[4] Bailer W, Thallinger G, Krawarik V, et al. AI for the media industry: application potential and automation levels[C]. International Conference on Multimedia Modeling. Cham: Springer International Publishing, 2022.

[5] Owsley C S, Greenwood K. Awareness and perception of artificial intelligence operationalized integration in news media industry and society [J]. AI & SOCIETY, 2024(1).

References:

[6] Zhang Hongzhong. "Artificial Intelligence + News" Leads the Trend of Media [J]. China Newspaper Industry, 2017 (11).

[7] Wang Xiaoming, Zhang Xiaohong. The Application and Development of Artificial Intelligence in the Field of News Communication [J]. Media Research, 2020(3).

[8] Khan M H. The Impact of AI on the Media Industry [J]. 2023.

[9] Lu Yongchun. Artificial Intelligence Promotes Media Transformation [J]. China Newspaper Industry, 2015 (23).

[10] Smith J, Johnson A. The Impact of Artificial Intelligence on Media Industry[C]. A Review. Journal of Media Studies, 2020(2).

[11] Chen L, & Wang Q. AI-Driven Personalization in Media Content Delivery: Opportunities and Challenges[C]. International Journal of Communication, 2019.

[12] Liu Y, & Zhang H. The Role of AI in Journalism: A Case Study of Automated News Writing[C]. Journal of Communication Technology, 2018(4).

[13] Li Hua, Liu Qiang. Research on Media Content Analysis Based on Big Data and Artificial Intelligence. Journalism and Communication Research [J]. 2018, 10(2).

[14] Zhang, Yu, Shengxiang Gao, Zhengtao Yu, Chengding Zhao, Peilian Zhao, and Peifu Han. "Case Element Joint Extraction Based on Case Field Correlation and Dependency Graph Convolutional Network." Neural Processing Letters 55, no. 5 (2023): 5583-5595.

[15] Wang L, & Smith R. Artificial Intelligence and the Future of Media: Opportunities and Challenges[C]. Journal of Media Innovation, 2019(3).

[16] Chen S, & Liu Q. The Role of Artificial Intelligence in Social Media Marketing: A Case Study of Brand Engagement[C]. International Journal of Advertising, 2018(4).

[17] Han, Peifu, Xue Li, Xun Wang, Shuang Wang, Changnan Gao, and Wenqi Chen. "Exploring the effects of drug, disease, and protein dependencies on biomedical named entity recognition: A comparative analysis." Frontiers in pharmacology 13 (2022): 1020759.