WEP Warwick Evans Publishing

International Journal of Social Sciences and Public Administration

ISSN: 3005-9836 (Print), ISSN: 3005-9585 (Online) | Volume 4, Number 2, Year 2024
DOI: https://doi.org/10.62051/ijsspa.v4n2.31
Journal homepage: https://wepub.org/index.php/IJSSPA/index

Mitigating Political Bias in AI-Generated News: A Comprehensive Analysis of NLP and Algorithmic Strategies

Yiming Sun[a, *], Zhuoer Lin[b]

RenMin University of China, Communication Studies, Sino-French institute, Suzhou, 215000, China

[a, *]2981902968@qq.com, [b]18896563836@163.com

ABSTRACT

The application of artificial intelligence technology in news production is becoming more and more widespread, and this paper investigates the problem of political bias in AI-generated news. Through the comprehensive use of natural language processing, machine learning and quantitative analysis, the news generated by different AI models at home and abroad are systematically detected and evaluated to quantify their political tendencies, and the political bias exhibited by different models in news generation and its specific forms are analysed. A set of solutions to reduce bias is ultimately proposed, with improvements in terms of dataset construction, algorithm design, transparency and accountability mechanisms, supervision and integration of multiple values, aiming to enhance the fairness and objectivity of AI-generated news.

KEYWORDS

AIGC (Generative Artificial Intelligence); Political Bias; Natural Language Processing; Quantitative Analysis; Algorithm Design; Dataset.

1. INTRODUCTION

1.1. Research Background

With the rapid development of information technology on a global scale, the application of artificial intelligence technology is becoming more and more widespread.AIGC technology uses natural language processing (NLP), machine learning and other methods to automatically generate news reports, which is gradually changing the production of traditional news and reshaping the pattern of information dissemination.

However, news content generated using AIGC technology may carry some degree of political inclination or emotional bias, which violates the basic principles of news reporting and is prone to possible adverse effects on public perception and social opinion. Therefore, it is necessary to understand and identify these biases and find out solutions to avoid them in order to enhance the transparency and credibility of news generation technology and promote the fair and objective dissemination of AIGC-generated news.

1.2. Literature Review

In exploring the application of generative artificial intelligence (AIGC) in the news industry and its potential problems, numerous scholars have conducted in-depth studies on this topic.The application of AIGC technology is gradually changing the way news is generated and distributed. For example,

Content from this work may be used under the terms of CC BY-NC 4.0 licence (https://creativecommons.org/licenses/by-nc/4.0/).
Published by Warwick Evans Publishing.

(Cai Zhiquan 2024) discusses the role of algorithms in news production, emphasizing the growing importance of automated systems in content selection and presentation[1]. However, the news content generated by AIGC is influenced by the algorithm design, training datasets and programming logic, which unconsciously reflects a specific political inclination or emotional colour.

Artificial intelligence technology not only has the inherent ideological attributes of science and technology in general, but also has new ideological attributes such as data hegemony and value shaping due to its technological nature of mimicking and replicating human intelligence[2] (Ye Geze. 2023).

Therefore it is very easy to create political bias. Political bias is a long-standing problem in news reporting, and it may manifest itself in many ways. With the development of AIGC technology, these biases may be unintentionally amplified or replicated by algorithms, thus affecting the fairness and objectivity of news content.

To detect and assess political bias in AIGC-generated news, researchers have proposed a variety of methods, including quantitative analysis and the application of natural language processing techniques. In addition, constructing a balanced dataset and optimising algorithm design are considered as key strategies to reduce bias[3]Xin Zhang. (2023). The introduction of transparency and accountability mechanisms and the combination of human oversight are also important measures to ensure the quality of news generated by AIGC[4] (Zhang, Linghan. 2019).

Through a comprehensive analysis of the existing literature, we can see that the application of AIGC technology in news generation is a complex issue that requires finding a balance between technological development and ethical responsibility. Future research needs to further explore how to optimise AIGC technology to reduce political bias and improve the quality and credibility of news content.

Based on this, this study proposes to explore and address the following questions: how can political bias in AIGC-generated news be systematically detected and assessed? How can existing methods be improved at the technical and ethical levels to reduce bias? By applying quantitative analysis, this study aims to propose a set of scientific and effective detection and assessment system, and to explore feasible ways to reduce bias in data collection, model training and result output.

1.3. Research Ideas

The first part focuses on the political tendencies and emotional colours of the news generated by different AI models at home and abroad, and uses a third-party platform to quantitatively score the degree of left- and right-leaning of each news item. The first part also combines the NLP model with linguistic analysis to explore the emotional colour and political stance of the news generated by different models behind the surface neutrality and the different phrases used.

Finally, combining the experimental results, we propose a set of systematic solutions aimed at overcoming the political bias implicit in the news content generated by AI models to the greatest extent possible, in order to improve the objectivity and fairness of AIGC-generated news. The significance of this research lies not only in the theoretical innovation, but also in the practice to promote the healthy development of AI technology in the field of journalism, and ultimately achieve a balance between technology and ethics.

2. ASSESSING AND DETECTING POLITICAL BIAS

2.1. Defining and Categorising Political Bias in News Reporting

To systematically detect and assess political bias in AI-generated news, a multi-level methodological framework is needed. We should accurately define bias and reasonably categorise it.

Definition of bias: The problem of bias in generative AI refers to the inconsistency of the results produced by the model for subjects of different countries, religions, races, genders, and other characteristics, and the latent risk of discrimination, ideological deviation, etc[5](Zihan Xiong, Chonghao Cheng, Jun Chen & Dabei Chen. 2024).

Political bias in news reporting usually refers to showing preference or hostility to a particular political position, party or policy in news reporting, thus affecting readers' perceptions and judgements.

Political bias can be manifested in a variety of ways, and common types include: selective reporting (the media intentionally or unintentionally report certain events or facts selectively, while ignoring or downplaying others.) , wording bias (the use of specific language or words to make one side appear more positive or more negative.) , source bias (selectively quoting or ignoring particular sources or experts to support a particular point of view.) , Emotional bias (conveying bias through emotional overtones in news reporting.) , agenda setting (influencing the public's perception of the importance of an issue through frequent coverage of that issue.) , framing bias (reporting the news through a particular frame or angle, thereby influencing the reader's understanding of events.) Picture and video bias (using specific pictures or videos to influence readers' perceptions and judgements.) , contextual bias (misleading readers' understanding of events by selectively providing or omitting contextual information.) etc. Some of the types of bias will be analysed in later sections with specific examples.

2.2. Generate News Stories and Evaluate Them

In this study, we use well-known domestic and international AI models such as chatgpt, claude, wenshin yiyin, kimi, etc., and let them generate news reports under the same premise of the same event instructions.

Based on the content of the news report, determine the political leaning of the news (left-wing, right-wing, neutral) and categorise it: Here the method of scoring with a third party AI model is used to quantitatively assess the news generated, scoring it on a scale of 0 to 10, with higher scores indicating stronger left-leaningness, and lower scores indicating stronger right-leaningness. a score of 5 represents complete neutrality. (See annex for input instructions and news content).

Table 1. Scores of news content generated by different big data models

	News I	News II	News III	News IV	News V	News 6	News 7	average score
Overseas I	5	6	7	5	7	5	5	5.71
Domestic I	6	6	7	5	7	5	5	5.57
Overseas II	5	5	6	5	5	5	5	5.14
Domestic II	5	6	6	5	5	5	5	5.28

Combined with the analysis of the different news scoring situations, it is clearly evident that with relatively vague and simple input instructions without emotional inclination colours, domestic and foreign AI-generated news is able to remain neutral and objective in attitude in most cases. In a few cases there is a political bias.

After carefully analysing the content of the news, it can be seen that a significant portion of the news has a more or less political bias especially present in the domestic AI, which generates news that is strongly influenced by Chinese ideological thinking. The following types of bias exist:

Frame bias: when it comes to the Taiwan issue, the news generated by domestic AI I emphasises that Taiwan independence is a counter-current, and ultimately settles on the fact that the two sides of the Taiwan Straits will be unified eventually; whereas the news generated by foreign gpt pays more attention to objectively analysing the realities in different fields in Taiwan, and settles on safeguarding Taiwan's interests.

Selective reporting: When generating news reports related to the situation of the Chinese stock market, although domestic AI I mentioned that the Chinese stock market was facing unstable factors, it emphasised at great length that the Chinese stock market was running steadily and in a good situation, and involved less in the problems encountered in the previous period. Domestic AI II, on the other hand, did not mention the challenges faced by the Chinese stock market at all. The news reports generated by foreign AIs clearly pointed out and analysed that China's stock market had experienced shocks and rising uncertainty, but the regulators had introduced various measures to make adjustments, combining the views of various parties and relatively objectively analysing the challenges and opportunities encountered by China's stock market.

Wording bias, emotional bias: When generating news reports on the Olympic medal table, the domestic AI I stood on China's position and called the United States a challenger and China's competitor. While the foreign AII objectively analysed the medal table situation and the competition situation of each country, switching the positions of each party, failing to see obvious political tendency and emotional colour.

It can be seen that some AIs are prone to selective reporting, wording bias, and emotional bias when generating news stories.

3. IMPROVING EXISTING METHODS TO REDUCE BIAS

The quality of generative AI generating and outputting content depends on the level of its database and model training, and its examination of the content depends on the feedback and domestication of human users. [6](Tang Zheng & Lin Zilu, 2023).

Therefore, reducing bias in AI-generated news is a complex and long-term process that needs to be thought about from both developer and user perspectives.

3.1. At the User Level

When providing instructions to the AI model, try to use clear and neutral language and avoid using words that may be biased or emotionally inclined. Provide instructions for multiple perspectives on the same event to allow the AI to generate reports from different angles to ensure comprehensiveness and diversity of news content.

a. Use of neutral, objective language

When providing instructions to the AI model, try to use clear and neutral language and avoid using words that may be biased or emotionally inclined. Provide instructions for multiple perspectives on the same event to allow the AI to generate reports from different angles to ensure comprehensiveness and diversity of news content.

Avoiding emotive words: Users should endeavour to avoid using emotive or evaluative words when writing instructions. For example, avoid negative words such as "extreme" and "bad".

Use specific descriptions: replace vague or tendentious terms with specific descriptions. For example, "the impact of government policies" is more neutral than "bad government policies".

b. Provide comprehensive and balanced background information

Multiple perspectives: Including multiple perspectives in the instructions ensures that the AI is able to generate comprehensive reports. For example, the instructions could include "Analyse the impact of the policy on different social" groups .

Detailed background: Provide background information in sufficient detail to avoid bias. For example, "Describe in detail the background to the policy and the views of the parties".

c. Avoid leading questions

Open-ended questions: use open-ended questions instead of leading questions. For example, "How do you evaluate the policy?" is more open than "Why is the policy unfavourable?" is more open-ended.

Neutrality questions: ensure that questions are worded neutrally and do not lead the AI to generate content in a particular direction.

Demonstration

Example a: Economic Policy Reporting

Bias Directive: Generate a story about how the government's new economic policies are hurting people.

Optimisation instructions: generate a report on the government's new economic policy, analysing its impact on different social groups and including different views from experts.

Example b: International Relations Coverage

Bias Directive: Generate a story about a country's hostile behaviour towards other countries.

Optimisation instruction: generate a story about a country's relations with other countries, containing positions and contextual analyses of both sides.

3.2. At the Developer Level.

Improvements are needed in a number of areas, including data sets, model design, transparency and accountability mechanisms, and supervision. By diversifying data sources, adopting advanced model design, increasing transparency and accountability mechanisms, and strengthening supervision, the problem of bias in AI-generated news can be effectively reduced, and the fairness and reliability of news reporting can be enhanced. The following are some specific strategies and methods:

a. Constructing balanced datasets

Build a balanced, comprehensive and diverse dataset. Ensure that the training dataset comes from a diverse range of sources covering different perspectives and positions to minimise the impact of a single viewpoint.

For example, data can be collected from news media from different countries, different political backgrounds and different cultural backgrounds. In the data pre-processing stage, obvious biased information is removed by cleaning and filtering techniques. Natural Language Processing (NLP) techniques can be used to identify and tag biased content and filter it before training. Avoid bias due to lack of data by ensuring that minority voices and perspectives are included in the dataset. This can be achieved through targeted collection and expansion of news coverage of minority groups.

b. Optimisation of algorithm design

Algorithm design should focus on fairness and diversity considerations.

For example, a multi-objective optimisation strategy can be used to take fairness, accuracy and diversity as one of the optimisation objectives; in addition, multi-modal information such as text, image and video can be combined to improve the comprehensiveness and accuracy of the model.

Multimodal learning can reduce the bias brought by a single modality and generate more objective news content through the fusion of information from multiple perspectives. Adversarial training methods are used to construct the generative model and the discriminative model to play with each other, prompting the generative model to continuously improve in reducing bias.

For example, a bias detector can be designed as a discriminative model to specifically identify and correct biased content in the generated news. Combining the strengths of multiple models, such as traditional rule-based approaches and modern deep learning methods, generates more balanced and unbiased news reports through hybrid models.

c. Increased transparency and accountability mechanisms

Interpretive AI technology is used to make the decision-making process of the model more transparent and allow users to understand the basis and logic of news generation. Interpretive AI can help identify and correct potential biases in the model and enhance users' trust in AI-generated news.

Encourage open data and modelling by allowing academia and industry to work together to review and improve AI news generation techniques. By opening up data and models, more experts can be attracted to participate in bias detection and correction, improving the impartiality of the models. Establish a clear accountability mechanism to regularly review and assess AI-generated news content. An independent review organisation can be set up to monitor and assess the impartiality of AI news generation and correct any bias problems found. Adhere to the consistency of rights and responsibilities, and in the process of AI research and development,application and management, the relevant entities or individuals must assume corresponding responsibilities and be held accountable for the consequences of the actions they generate[7] (Hu Feng & Zhang Chao.,2024).

d. Strengthening oversight

While AI news generation aims to be automated and intelligent, human oversight and intervention is still necessary. Human review should be introduced at key points to ensure the accuracy and fairness of generated news. While human review cannot completely eliminate bias, it serves as the last line of defence to identify and correct obviously biased content. In addition, a user feedback mechanism can be established to encourage readers to report and correct bias issues in AI-generated news. Through user feedback, the performance of the model can be continuously improved and biased content can be corrected in a timely manner. AI-generated news content should be continuously monitored and evaluated to identify and analyse the sources and manifestations of bias problems. Regular evaluation allows for understanding the effectiveness of model improvements and the development of further improvement measures.

e. Integration of diverse values

In the process of AI news generation, attention should be paid to the integration of multiple values. This includes giving full consideration to the views and needs of different groups, cultures and positions in data collection, algorithm design and news reporting; at the same time, multi-perspective and multi-dimensional news reporting is achieved through technological means, so that the public can get a full picture of the whole picture and the truth of the event.

4. CONCLUSION

This paper provides an in-depth study of the problem of political bias in AI-generated news. The results of the study show that although the generated news can remain neutral and objective in attitude in most cases, some AI models still show obvious political bias in specific contexts. These biases are mainly reflected in selective reporting, wording bias, emotional bias, etc. Especially when it comes to sensitive political issues, the domestic models are more influenced by their own country's ideology. Through quantitative analyses, we found that the political bias of generated news is affected by multiple factors such as dataset source, algorithm design, and input instructions. This study not only

proposes a theoretical framework for detecting and reducing the political bias of AIGC-generated news, but also provides guidance on how to build a fair and objective news generation system in practice.

Although this study proposes a series of strategies to reduce bias, such as improving input instructions, constructing balanced datasets, multi-objective optimisation strategies and transparency mechanisms, there are still challenges to fully implement these strategies in practice. For example, constructing a fully balanced dataset is difficult in practice, and the integration of multiple values in different cultural contexts requires more exploration. Future research can further refine the bias detection algorithm, explore ways to improve news generation technology in cross-cultural contexts, and validate and optimise the strategies proposed in this paper in more application scenarios.

REFERENCES

[1] Cai Zhiquan. Value Evidence, Technological Pitfalls, and Strategic Responses to Generative Artificial Intel ligence Enabling New Quality Productivity. Southwest Finance 1-14.

[2] Ye Geze. (2023) . Diverse Representations of Mainstream Ideological Risks in the Age of Artificial Intelligen Resolution. Journal of Theory (06) , 62-69 . doi:10 . 14110/j. cnki.cn-37-1059/d.2023 .06 .009.

[3] ZhangXin. . (2023) . Data Risks and Governance Paths of Generative Artificial Intelligence. Legal Science (Jou rnal of Northwest University of Politics and Law) (05) , 42-54 . doi:10 . 16290/j. cnki. 1674-5205 .2023 .05 .006.

[4] Zhang Linghan. (2019) . The rise of algorithmic power, alienation, and legal regulation. Legal Business Resea rch (04) , 63-75. doi:10 . 16390/j. cnki. issn1672-0393 .2019.04.006.

[5] Zihan Xiong, Chonghao Cheng, Jun Chen & Dabei Chen. 2011 A new genus and species of the genus Pse udourostyla (Coleoptera , Staphylinidae , Staphylininae) from China. (2024) . Security risks facing generative artificial intelligence . Communications Enterprise Management (06) , 54-57.

[6] Tang Zheng & Lin Zilu. (2023) . Generative artificial intelligence and journalism: empowerment, risks, and fore sight. Journalism & Writing (11) , 97-104.

[7] Hu Feng & Zhang Chao. Theoretical Review, Risk Revealing and Governance Path of Generative Artificial Intell igence - A Perspective of ChatGPT. Journal of Nanjing University of Posts and Telecommunications (Social Science Edition) 1 -11 . doi:10 . 14132/j. cnki. nysk.20240625.001.