

The Investigation of Political Bias in Large Language Model

Jiayi Lyu

Global Education School, Guangdong, China

qyylyx11@outlook.com

Abstract. Language models have deeply penetrated various social fields, including education, healthcare, and finance. With the widespread penetration and application of language models in these fields, and with public interest in their applications, the political bias in seemingly objective language model outputs is gradually attracting attention. Model political bias can not only interfere with the objective dissemination of information but also affect user cognition, public opinion, and even global interaction. This paper explores the causes of model language political bias from three dimensions: the data layer, the model layer, and the human layer. Data source bias, uneven geographical distribution, and text selectivity can contribute to data-level language political bias. Model optimization objectives imbued with political logic, as can correction mechanisms during model training and the selectivity of the model's technical architecture. Developer intent and manager needs can also contribute to human-level language political bias. Furthermore, this paper explores the specific manifestations of language model political bias: the issue stance, political entity descriptions, and sensitive topic handling in language model output. This paper analyzes the impact of language model political bias on users, society, and the world, as well as solutions to the problem. It is hoped that this article will provide a reference for understanding the nature of language model political bias among college students and the general public, promoting rational and objective use and the healthy development of language models.

Keywords: Big Language Model, Political Bias, Cause Analysis, Impact Assessment, Governance Paths.

1. Introduction

Big language models are developing rapidly and are closely related to people's learning, life, and work. Big language models are behind intelligent customer service answering questions, learning software helping students with homework, and mobile apps pushing information. Data recently released by the Ministry of Industry and Information Technology shows that my country's big language model market is expected to grow by 110% in 2024, with the market size further expanding [1].

Big model languages provide convenience for people to obtain the information they need, but they are not objective or neutral, and their political biases have a range of impacts on users. As college students are in the process of developing their personalities and cognition, the information conveyed by the big-model language they use daily may influence their political cognition. More broadly, the political bias of big-model language can also influence the formation of social and political consensus and impact national political security. Therefore, studying the political bias of big-model language has important political and practical significance [2].

Currently, academic research has begun on related issues. Xu Zeshui et al. analyzed the problems and future developments arising from big-language model-driven social development management, exploring the changes and issues emerging in social development management [3]. Based on relevant theories of risk society, this paper analyzes the risks and adaptation strategies that may arise from big-language model-driven digital government applications [4]. However, research on the political bias of big-model language is relatively rare. Existing research mostly discusses the potential political bias of big-model language from the perspectives of technical management and application scenarios, and further exploration is needed [5].

At the theoretical level, this study draws on Cheng Yulu's theory of the ideological transcendence of capitalist civilization by the new form of human civilization and Zhao Zhongyuan's analysis of the

Party's institutionalized ideological mechanisms to provide a theoretical basis for the generation and impact of the political bias of big-model language [1, 6]. Furthermore, by integrating relevant communication theories and drawing on the research of scholars such as Zhang Aijun on negative communication bias and online political security risks, this study explores the transmission path, scope of influence, and mode of action of the political bias of big-model language [7].

Big-model language is a language model trained on big data. It is an important branch of artificial intelligence, capable of understanding and generating human language and enabling human-computer interaction. For example, big-model languages, such as chatbots and writing assistants, are widely used in information query, content creation, and human-computer interaction [8].

Political Bias: In this study, political bias refers to the preference or tendency expressed by big-model language toward a particular type of political information during information dissemination, primarily encompassing language, political information retrieval, and political opinion guidance [2]. This study primarily draws on the following three theoretical perspectives: First, ideological theory analyzes the ideological sources of the political bias of big-model language and describes how the political will of different classes is reflected in big-model language. Second, drawing on the power construction of mainstream ideology disseminated in digital intelligence spaces by Song Lai et al., this paper explores an ideological analysis framework for big-model language. Third, this paper analyzes the power effects of big-model language ideology [9]. Second, communication theory explores the methods, scope, and impact pathways of the dissemination of big-model language's political tendencies, such as dissemination channels. This paper applies existing empirical methods such as Zhou Baohua's on the relationship between language models and public opinion [10]. Third, risk society theory draws on Lü Pei'an's analysis of the risks of applying big-language models to digital government and explores the potential political risks associated with the political bias of big-model language [4].

2. Analysis of the Political Causes of Bias

2.1. Embedding Political Attributes at the Data Level

2.1.1 Political Bias of Data Sources

The training data for big-models consists of data from various websites, news, and books, but these data sources harbor varying political ideologies. Some Western media sources are closely tied to the political systems and ideologies of their respective countries, and may be biased by their own political ideology when reporting on international politics. If these data samples constitute more than 60% of the training data for a large-scale model, the political ideology of the model-generated political content will be biased towards Western politics [2]. "If a student is exposed to only this data in junior college, using this data to train a large-scale model is likely to be influenced by a particular political ideology."

2.1.2 Data Imbalance from Regional Political Perspectives

Political data from different political perspectives may be difficult to fully train and cover in the training data for a large-scale model. This makes it difficult to encompass the political perspectives and data of all regions, potentially resulting in missing or unevenly distributed data on political topics in different regions. For example, a domestic large-scale model is primarily trained on Chinese data, and the proportion of political data related to political topics in the training data may vary significantly across different regions of the country. For example, compared to policy topics related to central and western China, the proportion of political data related to policy topics in parts of eastern China in the model training data may be over 30% higher. This may affect the comprehensiveness and accuracy of the model's processing and reflection of political topics in different regions of China [11].

2.1.3 Selective Collection of Political Event Texts

The political event texts used to train large models are somewhat selective. If the person collecting political event texts has their own political stance, political interests, or political environment, they will prioritize collecting political event texts or comments that support their own political stance and interests, while downplaying or ignoring those that do not. For example, when collecting text about a specific domestic political event, there is a tendency to collect text and comments that support the official political stance, while downplaying or ignoring text and comments that support neutral or opposing political positions. Some scholars have found that in some domestic political event training texts, text supporting the official political stance accounts for over 80%, while text opposing or neutral positions accounts for less than 20%. This selective collection directly leads to a strong bias in the model's processing of these political topics [2].

2.2. Internalization of Political Logic at the Model Level

2.2.1 Political Implications of Algorithmic Optimization Goals

The algorithmic metrics of large models are objective, but they harbor implicit political logic. Algorithmic optimization goals primarily focus on user experience and usage duration, but these goals also harbor some underlying political logic. For example, some platforms' algorithmic metrics are politically driven and conform to political needs. Some platforms' algorithmic metrics intentionally avoid or ignore the recommendation and push of sensitive political content, actively pushing content that aligns with national, political, and government needs [12]. The optimization goals set by these algorithms for recommendation results bias the model's expression of political discourse and the flow of political information, easily leading to an imbalance in the diverse political information exposed to students in vocational colleges.

2.2.2 Political Correction Mechanisms in Model Training

During model training, relevant personnel will correct habitual expressions in model output that do not conform to political requirements, retraining model content that does not conform to political requirements to conform to them. By repeatedly training the model on customary expressions that don't conform to political requirements, it establishes customary expressions that do. For example, the trained model corrects responses regarding national sovereignty and territorial integrity to ensure that the model's output conforms to national political requirements [13]. This mechanism can, while ensuring the compliance of political content, weaken the model's ability to express diverse political viewpoints, imbuing its language with a certain political bias.

2.2.3 Political Screening Features of the Technical Architecture

The technical architecture design of large models also inherently incorporates political screening. For example, text processing techniques such as political vocabulary screening and semantic analysis screen text for political leanings, prioritizing the retention and generation of text content that conforms to political requirements while removing, weakening, or adjusting content that doesn't. Technical documentation indicates that one large model has over a thousand rules for screening political keywords and phrases, blocking text containing these keywords and phrases. This inevitably results in the model generating content with a specific political bias [13].

2.3. Political Will Transmission at the Human Level

The human factor is one of the key factors affecting the accurate political reflection of large-scale model language. This influences the political nature of large-scale models through two factors. First, the varying experiences and perspectives of developers within the development team can easily imbue large-scale models with their own political will during design, training, and optimization. For example, core development personnel with their own political perspectives can easily bias their political orientation when setting training objectives, data, and parameter algorithms based on political principles [14]. Large-scale model development and management teams incorporate political

content into large-scale models based on national political requirements, guidelines, and policies. The management team supervises, inspects, monitors, and controls the large model based on political development requirements, and requires the development team to modify the model content to comply with political requirements. For example, if the model content involves the country's political system and must comply with political requirements, the development team will modify the model to comply with political requirements [9]. Table 1 provides three-level cause analysis comparison table of the large model's language political bias.

Table 1. Three-level cause analysis comparison table of the large model's language political bias

Causal level	Core influencing factors	Specific manifestation cases	Direct impact on junior college students
Data level	Data source position, regional data imbalance, and selective text inclusion	In the international model, Western media political data accounts for over 60%; in the domestic model, political data in the eastern region is 30%+ more than in the central and western regions.	Exposure to a single political perspective leads to a one-sided political understanding.
Model layer	Algorithmic political implications, political correction mechanisms, and technical architecture screening	Social platforms prioritize content with mainstream political values; one model sets filtering rules for over 1,000 political keywords.	Access to political information is limited, making it difficult to access diverse political viewpoints.
Human layer	Development team stance, regulatory requirements	Core development members tend to choose solutions that align with their own political views; regulators require model outputs to conform to official political statements.	Exposure to political information is the result of multiple political wills.

3. Analysis of Biased Political Manifestations

3.1. Political Inclination of Issue Stances

Large models exhibit distinct positions when discussing different political issues. On key domestic political issues, some large models consistently uphold sovereignty and territorial integrity and correct political values. However, when international political issues become hot topics, influenced by factors such as the dataset training level and the developer's background, some large models exhibit a distinct bias in their discussion of international political issues. For example, when discussing international political issues such as the Sino-US trade dispute, some Western large models mention the United States significantly more often than China when discussing political positions and demands, with the difference potentially exceeding 60%, which can easily lead to biased perspectives in students.

3.2. Bias in the Description of Political Entities

Large models tend to describe different political actors based on the criteria of "closeness" and "distantness," and emphasize their own political achievements when discussing political activities involving domestic political actors. For example, when describing the Chinese government's

achievements, they emphasize the nearly 100 million rural poor people lifted out of poverty and the complete removal of all 832 impoverished counties from poverty [15].

Some large-scale models exhibit political bias in their descriptions of political policies regarding political entities with which their own country is in political conflict or competition with other countries. When introducing a country's policies, they often focus on the political drawbacks and avoid positive political content. This prevents students from forming an objective and comprehensive understanding of that country's political entity and from developing a comprehensive and positive perspective on international politics [16].

3.3. Political Bias in the Treatment of Sensitive Topics

Some large-scale models adopt an evasive approach when faced with politically sensitive issues. First, they avoid the key issues in answering politically sensitive questions, providing only ambiguous answers and failing to provide useful political information. For example, when students ask about politically sensitive issues related to a specific period, some large-scale models simply respond with "This topic involves sensitive areas and it is hard to answer." This prevents students from understanding the relevant political history and society and from forming a comprehensive understanding of political history and society [14]. Some large-scale models lack objectivity in their approach to political issues. For example, some large-scale models are biased in their assessments of the advantages and disadvantages of different social systems, highlighting only one perspective. Some students have less knowledge and experience in politics and are easily influenced by relevant information [17]. Table 2 provides three Manifestations of Political Bias in Large-Scale Model Language.

Table 2. Three Manifestations of Political Bias in Large-Scale Model Language

Manifestation Type	Domestic Issues	International Issues	Cognitive Impact on Vocational Students
Issue Stance Bias	Firmly Safeguarding National Sovereignty	In Western Models, American Political Viewpoints Are Mentioned 40% More Often Than Chinese on the Issue of Sino-US Trade Friction	Biased Perception of International Political Issues
Political Entity Representation	Positively Expounding on the Political Contributions of the Own Government	Emphasizing the Negative Impacts of Policies on Political Entities in Competing Countries	Inability to Objectively Perceive Political Entities in Different Countries
Handling Sensitive Topics	Avoiding Sensitive Domestic Political and Historical Issues by Saying "It's Inconvenient to Answer"	Some Models One-Sidedly Emphasize Political Advantages and Disadvantages When Discussing Social Systems	Difficulty in Comprehensively Understanding Political History and Easily Misled

4. Governance Strategies and Outlook for the Political Dimension

4.1. Multi-Agent Collaborative Political Governance Path

4.1.1 Government Level

The government should exercise political oversight and leadership in the big model sector through the issuance of political documents, the formulation of standards, and policies [13]. First, further refine the political principles used in data collection for big model training, requiring that political data be diverse and objective. Second, strengthen the political content review mechanism for model outputs to ensure that the outputs align with national political requirements and the vital interests of the people. Third, strengthen the R&D of localized big models and reduce reliance on models with political biases [18].

4.1.2 Enterprise Level

Enterprises developing big models should be politically aware and sensitive, and pay attention to the political biases in model development [19]. The basic framework for analyzing the potential political biases of algorithmic models through the "data-algorithm-decision" process. The study suggests that the scope of political data acquisition should be expanded as much as possible, with emphasis on comprehensiveness and balance, to avoid political bias caused by the singularity of political data. Algorithm design should be improved to reduce the inherent political bias and avoid political bias caused by inappropriate algorithmic goal setting. Algorithm model design companies should establish a political ethics review mechanism and regularly review the political content output by trained models to reduce the political bias in their models [20].

4.1.3 School Level

Schools should provide specialized training in political perspectives and media literacy to students, enabling them to distinguish the political bias of large-scale model language [15]. Courses should introduce relevant knowledge about the political bias in large-scale model language, explaining the causes, expressions, and impacts of political bias in large-scale model language. This will help students realize that language is not objective and neutral, and that they should view political information objectively. Through political lectures and media literacy training, students can be taught to distinguish the authenticity and objectivity of political information, fostering independent political thinking and judgment, and reducing the risk of being misled by politically biased information [12].

4.1.4 Personal Level

Vocational students need to develop their ability to think independently and reduce their preference for large models [7]. "When faced with political information, maintain a skeptical attitude and avoid blindly following or accepting everything as it comes." "Verify the authenticity of political information from multiple sources, using methods such as corroboration, verification, and comparison through official websites such as the Chinese government website and Xinhua News Agency, political books, and other authoritative media to "cross-certify." "Actively participate in school political learning activities and media literacy training activities to "recharge" political knowledge and improve information discernment, "stay away from the political information cocoon," and establish a correct and comprehensive political perspective [21]."

4.2. Research Deficiencies and Outlook on the Political Dimension

4.2.1 Research Deficiencies

This study only discusses the expression biases of large-scale model language from the perspective of junior college students. However, as this is only a preliminary exploration of this issue, its breadth and depth are insufficient. In terms of data, due to the lack of large-scale research conditions, this article does not investigate and discuss the frequency of use of large-scale model language, common types, and scenarios in which political information is used. It cites already published data and research results, making it difficult to accurately reflect the extent of students' influence [10].

First, the research only selected a subset of mainstream domestic and international big models, making the sample unrepresentative and difficult to reflect the political bias characteristics of big models of different types and technologies. Second, the impact of factors such as model technology iteration and algorithm optimization on the political bias of big models was not included, and the objectivity and foresight of the research results need to be further improved [19].

4.2.2 Political Outlook

In the short term, with the further improvement of the government's political regulatory system and the continued attention paid by businesses to political ethics, the difficulty of managing the political bias of big model language will increase further. It is expected that in the next 1-2 years, my country will issue more detailed political review standards for big model training data, providing clearer guidance on the requirements and principles for entering big model training data. Businesses will also exercise closer dynamic control and management over the political output of big models to reduce unreasonable political bias, thereby ensuring that students and other users in a given discipline have more balanced access to specific political professional knowledge and information [18].

Table 3. Responsibility Table for Governance of Political Bias in Large-Scale Model Languages

Governance Subjects	Core Governance Measures	Specific Action Directions	Safeguarding Vocational Students
Government	Formulate political laws and regulations, conduct political oversight and review, and support local R&D	Clarify political data collection standards; establish a mechanism for reviewing political output content; increase investment in local model technology	Maintain political information security and reduce reliance on foreign politically biased models
Enterprises	Balance political data sources, optimize algorithmic political neutrality, and establish political ethics review	Broaden political data channels to avoid regional imbalances; reduce algorithmic political implications; regularly review model political outputs	Access more objective political model information to reduce cognitive misinformation
Schools	Political Education and Media Literacy Training	Integrate knowledge of model political bias into the curriculum; conduct lectures on discerning political information	Improve the ability to identify political bias and cultivate independent political judgment
Individual (vocational college students)	Verify political information through multiple channels, participate in political learning activities, and proactively question it	Verify model information using official platforms and professional political textbooks; participate in media literacy training; question politically biased content	Actively avoid political misinformation and avoid falling into a political cognitive cocoon

In the future, research on the political bias of large-scale model language can be conducted at a deeper level and in a wider range of fields. Researchers can explore more scientific and simpler

indicators for evaluating political bias and explore the impact of model political bias on specific groups, such as college students and teenagers, by combining their learning characteristics [20]. The continuous development of domestic large-scale model technology can better accommodate and balance diverse political viewpoints based on national political needs. While conveying mainstream political information, it can reasonably reflect diverse political viewpoints and enable students to acquire a more comprehensive political perspective [19].

Large-scale models will be further integrated with education. In the future, more large-scale models targeting students will emerge. In addition to assisting students in learning political knowledge and helping them to rationally interpret political information, they can also guide students to rationally view political information and identify political tendencies, becoming a powerful tool for cultivating students' political perspectives and discernment [15]. Table 3 provides responsibility Table for Governance of Political Bias in Large-Scale Model Languages.

5. Conclusion

Explore the causes of political bias in large language models from the data, model, and human layers. The former includes the political stance of the data source, regional political perspective bias, and the selective inclusion of political event text. The latter includes the political metaphors of the model algorithm, political correction mechanisms, and technical political screening. The latter also includes the political stance of the development team and the political requirements of management. Political bias primarily manifests itself in stances on political issues, representations of political entities, and the handling of politically sensitive topics. It impacts individual political cognition, the sociopolitical ecosystem, and international political interactions, exacerbating the cocooning effect of individual political cognition, exacerbating sociopolitical divisions and disputes, and impacting the balance of international political power.

To address these issues, this paper proposes a political governance approach encompassing four pillars: government, enterprises, schools, and individuals. Specifically, the government should strengthen political oversight and guidance; enterprises should strengthen political and social responsibility; schools should strengthen political education and media literacy; and individuals should improve their ability to discern political information. This paper also criticizes shortcomings in future efforts to strengthen political data research and research coverage, and offers perspectives on the management and research of political bias in big model language. Overall, this study offers a useful attempt to understand and manage the political bias in big model language, providing valuable theoretical and practical insights for the development of big models and their proper use by users.

References

- [1] 13.23 Billion Yuan! MIIT: China's Language Model Market Growth Rate Expected to Reach 110% This Year. *Intell Build Smart City*. 2024; (01): 5.
- [2] Huang YF, Huang LJ, Zhao M, et al. A study on the political bias of Taiwan media: A case study of cross-strait relations. *Taiwan Stud*. 2021; (06): 53-60. doi: 10.13818/j.2021.06.006.
- [3] Xu ZS, Sun TY, Zhu Z, et al. Management challenges and future trends of social development enabled by big language models. *Chin Manag Sci*. 2025; 1-19 [Epub ahead of print 2025-09-07]. doi: 10.1276/2024.1276.
- [4] Lv PA. Risks and adaptation paths of big language model application in digital government: from the perspective of risk society theory. *Hunan Soc Sci*. 2025; (04): 87-96.
- [5] Liu XB, Hu BT, Chen KH, et al. Key technologies and future development directions of large models: starting from ChatGPT. *Natl Sci Found China*. 2023; 37 (05): 758-66. doi: 10.16262/j.20231026.004.
- [6] Zhao ZY, Xue J. The internal logic and practical approach of the Party's ideological institutionalization. *Spec Zone Pract Theory*. 2025; 1-10 [Epub ahead of print 2025-09-07]. doi: 10.20250902.001.

- [7] Zhang AJ, Zong YN. Negative bias in communication: identification and mitigation of network political security risks. *J Harbin Inst Technol (Soc Sci Ed)*. 2024; 26 (02): 34-42. doi: 10.16822/j.2024.02.003.
- [8] He Z. An application of intensive intelligent knowledge management based on large language models. *Wireless Internet Technol*. 2024; 21 (09): 46-8.
- [9] Song L, Wang HJ. From diaspora to bridging: the construction of the power of mainstream ideology in digital intelligence space. *Leadership Sci*. 2025; (05): 152-60. doi: 10.19572/j.2025.05.012.
- [10] Zhou BH, Fang Y. Can large language models reflect public opinion? An empirical study based on a three-dimensional framework of data resources, opinion distribution and prompt language. *Journal Commun Res*. 2025; 32 (05): 46-64,127.
- [11] Zhang L, Pi JY, Song GX. Local government competition and productive expenditure bias: a political economic analysis of abolishing counties and establishing districts. *Finance Trade Econ*. 2018; 39 (03): 65-78. doi: 10.19795/j.2018.03.006.
- [12] Ren YN, Li XQ. Value logic and risk governance of digital transformation of ideological and political education enabled by algorithm recommendation. *J North Univ China (Soc Sci Ed)*. 2025; 1-9 [Epub ahead of print 2025-09-07]. doi: 10.20250828.1640.006.
- [13] Bao CF. Ideas for building a security assurance system for the application of large conversational language models. *China Inf Secur*. 2023; (05): 55-9.
- [14] Zhang AJ, Yang DN. The development results and governance strategies of secondary political opinion in the era of generative artificial intelligence. *Exploration*. 2025; (04): 149-60. doi: 10.16501/j.2025.04.011.
- [15] Li Y, Liu ZL. Research on the effectiveness of governance in precisely providing ideological and political education in colleges and universities in the digital age. *Party Build Ideol Educ Sch*. 2025; (14): 57-60. doi: 10.19865/j.2025.14.013.
- [16] Zhang TY. The duet of political bias and media bias: an exploration of the irrational tendency of people in the Northeast Asian community to construct each other's national images: based on an investigation centered on China and Japan. *Journalism*. 2017; (09): 2-7. doi: 10.15897/j.2017.09.001.
- [17] Tan CY, Liao HL. New media and socialist democracy: technological empowerment and communication bias. *J Yunnan Admin Coll*. 2018; 20 (06): 75-9. doi: 10.16273/j.2018.06.041.
- [18] Li KK, Li ZH. Domestic large models usher in a tier-1 competition landscape. *China Bus News*. 2025 Apr 14; C04. doi: 10.38300/n.2025.000760.
- [19] Xiao Q. Research on the technology and application of large language models for government vertical fields [dissertation]. Chengdu: Univ Electron Sci Technol China; 2025. doi: 10.27005/d.2025.001914.
- [20] Li X, Zhu R, Chen XL, et al. A review of hallucinations of large visual language models: causes, evaluation and governance. *Comput Res Dev*. 2025; 1-24 [Epub ahead of print 2025-09-07]. doi: 10.20250506.1509.006.
- [21] Zhu RH. Research on the governance path of public relations algorithms for online ideological and political education. *Public Relat World*. 2025; (14): 181-3.