

Research on the Application Potential, Risks and Challenges of Generative AI in University Ideological and Political Education and Ethical Regulation

Xiaoyu Shang, Shuairan Tang, Yijie Xin

Sanya University, Sanya, 572022, China

Abstract: Generative artificial intelligence (GenAI) technology offers innovative opportunities for ideological and political education in colleges and universities but also presents multiple challenges. This paper discusses the potential of GenAI to empower ideological and political education by serving as an intelligent dialogue partner, content generator, and personalized tutor, as well as by innovating models, supplying resources, and providing personalized teaching. The paper analyzes the risks GenAI poses to ideological security, academic ethics, data privacy, and algorithmic responsibility. It proposes building a comprehensive ethical regulatory framework with value guidance as the core, multiple collaborative governance as the framework, literacy improvement as the basis, and technological empowerment as the guarantee to promote the healthy and orderly development of GenAI in ideological and political education. This framework would serve the fundamental goal of establishing morality and cultivating people.

Keywords: Generative Artificial Intelligence, Ideological and Political Education, Ethical Risk, Ideological Security, Governance Strategy

1. Introduction

Generative artificial intelligence (GenAI), represented by ChatGPT and DeepSeek, is causing a global industrial and social transformation due to its strong content generation, context interpretation, and natural interaction capabilities. GenAI has enabled a wave of technological diffusion in many fields, including media, finance, and medical care. Education is regarded as one of the most promising GenAI application fields due to its vast knowledge base and the nature of dialogue and interaction. Compared to general knowledge education, university ideological and political education plays an important role in moral education. It is not only related to teaching students knowledge, but also to guiding their outlook on life, shaping their beliefs, and ensuring ideological security. Therefore, GenAI's intervention is far from simple technological empowerment. The opportunities and challenges it brings are more profound and complex and urgently need careful and systematic examination.

Currently, academic and practical circles have preliminarily explored the application of GenAI in education. However, the results are generally "dual-track parallel" yet "biased." On the one hand, GenAI provides a large number of new methods for ideological and political education and endows it with significant application potential, such as realizing targeted learning as an "intelligent tutor" and innovating teaching resources as a "content generator." This shows a tendency of "technological optimism"^[1] (Lyu, 2024). On the other hand, some scholars have warned of its potential risks, especially the crisis of academic integrity, algorithmic bias, and other ethical issues it may cause^[2] (Miao, 2025). Therefore, it is crucial to address the "double-edged sword" effect that GenAI has on the ideological and political education of colleges and universities. Only by making full use of its innovative potential, preventing and resolving its multiple risks, and ensuring ideological safety can we promote the integration of ideological and political education with artificial intelligence ethics and enrich the ideology of the technological era of the new century. This will guide the healthy, orderly, and positive development of GenAI in the field of ideological and political education.

2. Empowerment and Innovation: The Application Potential of Generative AI in Ideological and Political Education

Generative AI (GenAI) provides a new technical path for overcoming some of the limitations of

traditional university ideological and political education due to its strong ability to process language and generate content. GenAI's empowering potential has invigorated the development and digital transformation of ideological and political education by innovating interaction modes, resource forms, and teaching and learning processes.

2.1. As an intelligent dialogue partner: Innovate the teacher-student interaction model.

Personalized and in-depth ideological guidance is difficult to achieve in large-scale college and university majors, which is a major challenge for traditional ideological and political education. GenAI can serve as intelligent dialogue partners that, to a certain extent, eliminate such challenges. First, GenAI can answer questions and supplement knowledge from thousands of people. Students can ask GenAI questions about any concept or historical event in ideological and political education at any time and receive immediate, accurate answers. GenAI helps students consolidate their theoretical foundation and helps teachers identify key issues, alleviating the pressure of after-class counseling. Second, GenAI has the potential for sentiment analysis and preliminary thinking. When students have emotional problems, GenAI can simulate empathetic dialogue to help them express their confusion and eliminate negative emotions. GenAI can then provide preliminary advice and guidance to students, such as helping them alleviate pressure related to school or interpersonal issues. This can also provide effective assistance for subsequent teacher intervention. However, it's important to recognize that this kind of interaction is essentially algorithm-driven and cannot replace the deep communication between teachers and students that involves true feelings and life experiences. It should be clear that GenAI's role is supplementary, not alternative.

2.2. As a massive content generator: Innovate the supply of teaching resources.

Ideological and political education requires content that is up-to-date and relevant to social reality. As a powerful "content engine," GenAI can aggregate public domain resources (Learning Power, Data China, etc.) and private domain resources (human-machine dialogue) to create an enriched collection. As a massive content generator: Innovate the supply of teaching resources. Ideological and political education requires teaching content that keeps up with the times and is close to social reality. As a powerful "content engine," GenAI can aggregate public domain resources (Learning Power, Data China, etc.) and private domain resources (human-machine dialogue) to create a collection that can enrich the form and efficiency of teaching resources. First, GenAI can dynamically generate highly contextualized teaching cases. Teachers only need to input keywords, such as "rural revitalization" or "green economy," and GenAI can quickly generate textbooks that are detailed and close to life, making practical teaching more vivid. Second, GenAI can simulate current affairs comments with diverse viewpoints. During class debates or case analyses, GenAI can simulate the views of individuals holding different positions. This helps students understand the composition of the complex public opinion field and trains their critical thinking and discrimination abilities. These capabilities aim to improve teaching efficiency. However, teachers still need to perform a final political check and academic review of the generated content. Finally, GenAI can integrate the public and private domains of a set of intelligence factors into ideology and culture, making them static and realizing creative transformation and innovation in terms of ideological development^[3] (Feng, Lin, & Guo, 2024).

2.3. As a personalized learning tutor, GenAI can facilitate individualized teaching.

In traditional teaching, the "many-to-one" model is often used, which prevents teachers from considering the listening comprehension and knowledge retention abilities of some students. However, GenAI can help teachers cultivate students' professional skills and display and integrate the ideological and political aspects of the curriculum. Based on the comprehensiveness of the information collected in the database and the professionalism with which the big data model is derived, GenAI can deduce students' personal status and degree of understanding according to the interrogation mode and time when facing the questioner. This allows the average value of students in a domain to be analyzed. To improve teaching efficiency, the teaching management mode is more targeted and professional. Second, GenAI is no longer bound by time and space limitations, breaking through traditional teaching norms and enabling students to explore independently during highly concentrated time periods. For introverted students, multidimensional interaction contributes to positive study and the search for knowledge, helping them overcome their fear of asking questions. Finally, as Tom Hooper, founder of Third Space Learning, puts it, "Teachers will not be replaced by AI, but rather, AI will support them to become better teachers." Ideological and political education maximizes the moral education function of the curriculum and distills

the cultural values contained within. In teaching, teachers use daily interactions as a foundation and GenAI as an effective communication tool to seamlessly integrate spiritual guidance.

3. Worries and Challenges: Analysis of the multiple risks brought by generative AI

Generative AI is assigned to innovation education and educational innovation. At the same time, its inherent logic and application paradigm have caused a lot of concern that cannot be ignored. These risks and challenges are not merely side effects of technology; they impact the core values and safety bottom line of ideological and political education at a deeper level. It is urgent to systematically analyze and respond to these risks and challenges. Ideological security risks: the prominence of core risks

3.1. Ideological security risks: the prominence of core risks

Ideological education is at the forefront of national ideological security. The "value" problem of emergent AI makes it a potential source of risk.

First, algorithmic bias and the implicit penetration of Western values are challenges that cannot be underestimated. Algorithmic bias can be divided into pre-existing algorithmic bias, technical algorithmic bias, and emergent algorithmic bias. Since the latter two cannot be effectively solved under normal circumstances, this paragraph will focus on antecedent algorithm bias. This refers to when algorithm designers intentionally or unintentionally embed their biases, institutional systems, and cultural differences into the algorithm during programming. Since GenAI model training data comes from global internet information, it is unavoidably embedded with data bias and value orientation^[4] (Liu, 2019). People subconsciously believe that such algorithms are objective and mostly ignore the hidden value choices behind them. When these models generate content for college students, it may unconsciously incorporate a Western conceptual framework, individualistic values, and implicit bias against theories and systems with Chinese characteristics. For instance, when answering questions about "democracy" and "human rights," GenAI may generate discourses that align with Western liberal discourse systems rather than providing comprehensive explanations based on China's national conditions. This kind of cultural value output under the false appearance of "technology neutrality" is deceptive and has a strong influence on shaping college students' socialist core values.

Secondly, the "information cocoon room" effect solidifies the information acquisition path and narrows the discourse power of multiple ideologies. GenAI's filtering mechanism and personalized push operation continuously provide users with content similar to their existing interests, beliefs, and positions. Students who rely on AI for information over the long term may have limited individual contact with the "filter bubble" constructed by the algorithm, which narrows search results, strengthens individual preferences, and weakens multiple information channels^[5] (Peng, 2020). This hinders the development of students' comprehensive and dialectical logical thinking abilities and makes heterogeneous information, brought by diverse individuals, fragmented and marginalized. This weakens the effect of introspective literacy education and the cohesion of value consensus aimed at ideological and political education.

3.2. Academic theory and educational value risks: The impact of direct risks

Ideological and political education involves teaching knowledge, training thinking abilities, and embodying the innovative spirit of contemporary students. However, the convenience of the "one-click generation" feature of generative AI hinders its educational purpose.

On the one hand, the emergence of generative AI has an impact on the academic integrity system. Directly copying and pasting generative AI text violates the core principles of academic integrity and fails to reflect the author's independent thinking and opinions, constituting a form of implicit academic deception. When faced with tasks such as final subject papers and practice reports, users only need to provide simple instructions to generate the desired content. However, it is difficult to distinguish the authenticity of content generated by AI, as it is with human-generated content. Subjective and objective evaluations will eventually call the authenticity of the article into question, making it difficult to reach a fair and just conclusion. The article's reference value will be criticized, and academic integrity will also be called into question.

On the other hand, critical thinking will increasingly weaken, breeding lazy thinking. In the past, anthropologists had to express their opinions through speeches, articles, or interviews with the media. People needed to read books and watch documentaries and news broadcasts to understand cutting-edge

academic research. However, generative AI breaks these traditional methods. It only needs to integrate all kinds of information through dialogue without understanding or thinking. This makes students develop technical skills. The core value of academic research lies in researchers' original contributions, critical analysis of ideas developed through brainstorming, and addition of a theoretical foundation for global education. College students use AI to summarize literature and analyze data, affecting the direction of research and eroding their ability to solve complex problems and think independently. This results in thinking inertia.

Additionally, the authority and subjectivity of education are deconstructed. One of the main goals of education is imparting knowledge. However, with the expansion of technology, the value of teachers has been weakened. Generative AI can only answer based on a pre-set knowledge base, so the teacher becomes the one being manipulated. In human-computer interaction, the boundary between educators and educational objects gradually blurs, causing students to become more dependent on and trusting of generative AI and eliminating the role of educators in guidance. As a result, it is difficult to achieve accurate ideological and political education in the college and university curriculum, and it deviates from the practical direction of ideological and political education in shaping the soul of education.

3.3. Ethical risks of data and algorithms: Theft of basic risks

Because many users jointly use generative AI and directly feed it data, GenAI continues to have data sources. The privacy of user information is questionable, as is the ownership of product responsibility.

First, the internet poses the hidden danger of data privacy leakage. The emergence of artificial intelligence raises ethical issues regarding the protection of digital users' privacy. Discrimination and prejudice involving algorithms subsequently create "information cocoon rooms." This involves the use of user-generated AI that recommends content to users based on data analysis. This presents a relatively "personal" digital world. The development of artificial intelligence requires a large amount of data. "Personal information processing is an important stage for the algorithm platform of personal privacy resources."^[6](Yang Xixi, Li Bengan, 2023) Although data is fed to AI anonymously, it deeply infringes on people's privacy, resulting in the continuous supply of personal information to digital platforms. If sensitive personal information held by intelligent systems is leaked, people may find themselves in embarrassing or even dangerous situations, and their privacy rights will be infringed upon. Users' privacy data is stolen consciously or unconsciously, which puts their personal information at risk.

Second, there is the dilemma of decision-making opacity and responsibility attribution caused by the algorithm black box. Referring to the data provided by the AI without rigorous verification is equivalent to using an unviewable "black box" during the experiment. This can lead to algorithmic discrimination in the results, which can have negative consequences. When research results are disseminated, who is responsible for serious errors: the authors, the developers of the AI tools, or the reviewers? "The algorithm is the basis for the development from information body to agent, and it is from this that the autonomous intelligent algorithm is gradually acquired"^[7](Wang Tian'en, 2024). Therefore, AI intelligence can only be used as a tool and has no authority. In other words, human authors should exercise critical thinking when using AI in their research, otherwise they will be held responsible for any mistakes. In processing massive data, the logic of artificial intelligence algorithms is difficult for people to understand, and it is impossible to grasp their operation logic. This creates an opaque orientation. This kind of "black box problem" cannot be traced to the source and brings ethical problems to AI that cannot be explained or predicted. In short, AI can only be used as a tool for ideological and political education to eliminate the negative effects of the "black box" problem and opacity.

4. Guidance and Regulation: Building a Governance Path for Future Ideological and Political Education

In the face of the various difficulties posed by GenAI, a passive response or simple technical control is not the best policy. From a Marxist philosophical perspective, technology and human beings are in a two-way interactive practical relationship of "subject objectification" and "object subjectification" because technology is the product of the objectification of humans' essential power ^[8](Li Meifeng & Li Yi, 2008). This means that we should neither regard technology as an objective tool nor fall into the misunderstanding of technological determinism. Instead, we should adhere to the basic principles of Marxism and actively build an organic whole supported by value guidance, subject-object coordination, quality improvement, technological empowerment, and other multidimensional aspects. This will allow us to achieve the deep integration and value identification of GenAI in the field of ideological and

political education. This will allow us to achieve the fundamental purpose of serving people, establishing morality, and cultivating individuals.

4.1. Value Guidance: Establish the ethical principle of "people-oriented, doing good."

The ultimate goal of technology governance should be to serve the comprehensive development of human beings. In the field of ideological and political education, which has a unique position, socialist core values must be the fundamental ethical guidance and the highest criterion. The entire life cycle of GenAI applications in this field should be strengthened. We should focus not only on the efficient use of technology but also on value guidance. Specifically, during model design, data screening, content generation, and result evaluation, we must remove the potential influence of Western centric ideology. We should establish the concept of a community with a shared future for mankind and a value review and ethical alignment mechanism based on fairness and justice. As relevant research points out, "the more complete the training data, the more accurate the algorithm." However, if the training data contains cultural and value biases, the algorithm's output will inevitably deviate from fairness and objectivity^[9](Wang & Ru, 2020). Therefore, GenAI's external output must fundamentally conform to China's mainstream ideology, ideological and political education, morality, and social values to achieve China's educational goals.

4.2. Multiple Coordination: Building a Multi-Level, Three-Dimensional Governance System

The governance model of a single entity is difficult to use to solve the complex problems brought by GenAI. Therefore, it is necessary to establish a top-down, bottom-up, interactive governance system.

At the national level, the focus should be on top-level design and drawing red lines. The government should play a leading role in formulating laws, regulations, and national standards for AI applications in education that clarify the requirements. They should also establish a filing and review system for GenAI educational applications and carry out strict value and safety assessments on AI products entering campuses to block risks at the source. China has made clear its commitment to "paying equal attention to development and safety, and parallel innovation and ethics" in AI governance. The country has outlined the system requirements for algorithm filing and safety assessments in the Regulations on the Management of Internet Information Service Algorithm Recommendation and the Measures for the Management of Generative Artificial Intelligence Service (Draft for Comments). These provide an important reference for the application and governance of AI in education^[10](Cao, 2023). Looking to the future, the GenAI governance framework must be more prudent, balancing innovation and security while controlling risks.

At the university level, we should improve school rules and strengthen responsibility. It should be made clear that colleges and universities bear primary responsibility for governance. Rules for the use of GenAI in universities should define the boundaries of AI use by teachers and students for teaching, research, and learning purposes, as well as the corresponding responsibilities and punishments for prohibited behaviors. For example, using AI to write papers and report assignments is clearly listed as academic misconduct, and the corresponding disciplinary measures should be taken. Honest people enhance confidence.

At the social level, there should be industry self-discipline and joint governance. Encourage all major AI industries to establish conventions and strengthen enterprise initiatives to conduct value, alignment, and security audits. Additionally, establish internal ethical review and cross-evaluation mechanisms. Meanwhile, independent third parties should evaluate the technical compliance and ethical standards of AI outputs. Additionally, the GenAI public reporting and feedback platform was developed to incorporate public supervision into the governance process.

Government supervision, university strengthening, and social supervision form a governance structure to guide the healthy development of GenAI.

4.3. Literacy improvement: Launch education guidelines on the use of AI for teachers and students

In addition to the synergy between society and values, teachers and students should have relevant action strategies. Colleges and universities should train ideological and academic planning tutors on GenAI usage norms. For example, they should learn how to collect information from an AI database. How does its mode of operation affect the output information? How can reasonable judgments be made about content generated by AI? Is the content and academic information accurate? Through relevant

skills training, teachers can transform from users of AI tools to facilitators of new-era products. For students undergoing changes in the age of diverse information, using GenAI is a "required course" in their academic careers. Students can learn relevant content systematically in courses such as information technology or ideological and political education. They can effectively avoid the "black box" effect caused by information blocking and deviation in value orientation. Students will develop critical thinking, tool operation, and review abilities. They will learn how to use various learning websites and judge the reliability of information sources independently. Academic integrity education is strengthened, and the boundaries and norms of AI-assisted learning are clarified to guide students in establishing AI ideological and political ethics and becoming literate digital citizens.

4.4. Technology Empowerment: Explore solutions that are compliant, controllable, and valuable to them

Before generative AI is developed and applied on a large scale, relevant government education departments should strengthen technical supervision to avoid system risks and vulnerabilities. Technology departments must protect teachers' and students' rights to know and control their data. They should clarify what data has been collected. How is it being used? The correction, access, or deletion of such data is at the discretion of the data provider. On April 11, 2023, the Cyberspace Administration of China publicly solicited a total of 21 comments on the Generative Artificial Intelligence Service Management Method (Draft). This document initially provided more detailed regulations for the generative artificial intelligence industry. The establishment of this method indicates that GenAI is subject to strict monitoring and management to protect the interests of all parties and ensure that educators and students adhere to relevant regulations, improving the data supervision system. To ensure the accuracy and value of artificial intelligence for innovation and research, relevant research departments must address data fairness and ethics. To strengthen legal education regarding virtual data, the United Nations Educational, Scientific, and Cultural Organization (UNESCO) stated in the Education 2030 Agenda that "UNESCO essentially requires a people-oriented approach to artificial intelligence." Ensure that AI does not widen the gap between individuals and nations. Ultimately, use generative AI-specific models or related security plugins to empower technology.

5. Conclusions and Prospects

The study found that generative AI can be a reliable conversation partner and personalized mentor on a variety of issues. However, due to its ideological penetration, the crisis of academic integrity, and data security risks, the application of generative AI is more stringent. It is difficult for audiences to distinguish truth from falsehood, resulting in a trust crisis of generative AI. This crisis can cause irreparable damage to personal reputation, companies, and even social stability. At the macro level, society and universities should establish severe punishment mechanisms, cross-platform linkages with government departments, and close cooperation with regulatory agencies and legal departments. They should also establish authoritative safety inspection mechanisms. At the meso level, students should be guided to develop critical thinking skills when receiving information. They should learn to identify the authenticity of content and use cross-verification to make rational judgments about the reliability of information. At the micro level, improving the digital literacy and ethical awareness of teachers and students is fundamental to mitigating risks. Strengthening content traceability and digital forensics is necessary to avoid risks and actively explore. For example, blockchain technology should be used to ensure immutability and improve the effectiveness of interpretive generative AI in ideological and political education. Forward-looking ethical regulations should be adopted to address risks in generative AI's ideological and political education.

This study provides a systematic review of research on the potential, risks, and ethical regulations of generative AI in ideological and political education at colleges and universities. However, due to insufficient databases, it is impossible to accurately collect large amounts of teacher behavior data and student value change data. There is also a lack of firsthand information for universities. Therefore, the conclusions are limited in availability and accuracy, and their persuasiveness needs to be strengthened. In the future, we will start from three working points. First, we will collaborate with computer science and Marxist theory courses to develop a specialized "AI ideological and political education" model and align the educational objective with the algorithm. Second, we will conduct substantive data collection and interviews to evaluate the influence of contemporary college students on their political identity, moral judgment, and digital literacy in real time. Third, we will conduct an international comparison between China and other European and American countries to analyze the differences in the educational effects of

generative AI in various institutional environments. This will provide support for the development of AI ideological and political education in China.

References

- [1] Lv Y Y. Application prospect of ChatGPT in University education evaluation: Logical evolution and development direction [J]. *Journal of yunnan normal university (philosophy and social sciences edition)*, 2024, 56 (03): 127-136. The DOI: CNKI: SUN: YNSF. 0.2024-03-014.
- [2] Miao F Y. Risk, cause and resolution: data manipulation, algorithm bias and ethical review of generative artificial intelligence from the perspective of contextualism [J]. *Journal of kunming university of science and technology (social science edition)*, 2025, 25 (02): 49-56, DOI:10.16112/j.carolcarrollnki.53-1160/c.2025.02.122.
- [3] Feng L, Ni G L. Digital transformation of ideological and political education based on generative artificial intelligence [J]. *Ideological education research*, 2024 (02): 46-53. DOI: CNKI: SUN: SIXI. 0.2024-02-007.
- [4] You-hua liu. And its regulation path algorithm bias [J]. *Journal of law*, 2019, 40 (6): 55-66. The DOI: 10.16092/j.carolcarrollnki.1001-618-x.2019.06.006.
- [5] Peng Lan. Multiple Factors Leading to the Information Bubble and the Path to "Break Free" [J]. *Journal of Journalism*, 2020, (01): 30-38 + 73. DOI: 10.15897/j.cnki.cn51-1046/g2.20191230.001.
- [6] Yang Xi-xi, LI Ben-qian. The generation of structural deprivation: algorithm audience privacy deprivation logic [J]. *Journal of shandong normal university (social science edition)*, 2023 of (6): 125-138. DOI: 10.16456/j.carolcarrollnki.1001-5973.2023.06.012.
- [7] Tian-en wang. The evolution of the artificial intelligence algorithm and its ethical effect [J]. *Journal of shanxi normal university (social science edition)*, 2024 ploydy (02): 36-44, DOI: 10.16207 / j.carolcarrollnki.1001-5957.20240206.001.
- [8] Elizabeth lee, Li Yi. The rationality of technical defense [J]. *Science, technology and dialectics*, 2008, (01): 66-70 + 112. DOI: CNKI: SUN: KXBZ. 0.2008-01-014.
- [9] Wang Huai-jun, RU Xu-hua. Artificial intelligence algorithm discrimination and its governance [J]. *Philosophy of science and technology research*, 2020 (02): 101-106. The DOI: CNKI: SUN: KXBZ. 0.2020-02-017.
- [10] Cao J F. Toward Trusted AI: Governance challenges and countermeasures of ChatGpt-like generative artificial intelligence [J]. *Journal of Shanghai institute of political science and law (review) under the rule of law*, 2023, 38 (04): 28-42. DOI: 10.19916/j.carolcarrollnki.cn31-2011/d. 2023.04.005.