

Algorithmic Disenchantment and Value Reshaping: Ethical Risks and Governance Mechanisms of Embedding Generative AI into University Ideological and Political Education

Yuchen Liu ¹, Feng Zhong ², Yingmei Li ^{3,*}

¹ School of National Finance, Guangdong University of Finance, Guangzhou, China

² School of Credit Management, Guangdong University of Finance, Guangzhou, China

³ Qingyuan Campus, Guangdong University of Finance, Guangzhou, China

ABSTRACT

As Generative Artificial Intelligence (Generative AI) becomes deeply embedded in the field of higher education, Ideological and Political Theory Courses (hereinafter referred to as "Civics Courses") in universities are undergoing a paradigmatic transformation from "digitization" to "intelligence." As a disruptive technology, Generative AI has greatly enhanced the precision and sense of gain in ideological and political education through automated content production, anthropomorphic emotional interaction, and personalized scenario construction. However, the strong intervention of technological logic has also triggered deep-seated ethical concerns: the algorithmic "black box" may obscure the light of truth, leading to "technological dependence" and "mental laziness" in cognition; data bias may implicitly dissolve the authority of mainstream ideology, triggering "algorithmic colonization" of values; and the alienation of human-machine relationships may lead to the dual withdrawal of teacher and student subjectivity, causing "emotional hollowing" in the educational process. Facing these challenges, it is imperative to perform "algorithmic disenchantment" on Generative AI, penetrating the capital logic and technological limitations behind its instrumental rationality. Furthermore, "value reshaping" in ideological and political education must be achieved through paths such as constructing ethical regulations for "human-machine synergy," creating independent and controllable vertical models, and reshaping the digital literacy of teachers and students. This paper aims to explore the dialectical unity of technological empowerment and value leadership from the three dimensions of ontology, epistemology, and axiology, providing an ethical barrier and governance scheme for the high-quality development of university Civics Courses in the intelligent era.

KEYWORDS

Generative AI; University Ideological and Political Education; Ethical Risks; Algorithmic Disenchantment; Value Reshaping; Governance Mechanism

1. INTRODUCTION: "TECHNOLOGICAL PRESCRIPTION" AND "ETHICAL SYMPTOMS" IN THE INTELLIGENT ERA

Marx once profoundly pointed out in *Das Kapital*: "Along with the tool, the skill of the workman in handling it passes over to the machine." [1] In the current era, the explosive growth of Generative AI technologies, represented by ChatGPT and Sora, is reshaping the modes of human knowledge production and rules of social interaction. As key courses for implementing the fundamental task of fostering character and civic virtue, university Civics Courses are inevitably swept into this wave of technological change.

With its powerful natural language processing capabilities, multimodal generation abilities, and human-like logical reasoning, Generative AI is regarded as a "technological prescription" to solve chronic problems in traditional Civics Courses, such as "low head-up rates" (student engagement), "weak pertinence," and "poor interaction." It can instantly generate massive teaching cases, construct immersive virtual historical scenarios, and even play the role of "Socrates" to conduct deep value debates with students. However, technology is not merely a tool but a type of "Enframing" (Gestell). When we introduce Generative AI into the Civics classroom, we also introduce the entire set of technological logic and algorithmic rules behind it.

Currently, academic research on AI-empowered ideological and political education mostly focuses on the instrumentalist level of "application landscapes" and "efficiency improvement," with insufficient attention paid to the "ethical symptoms" brought about by technological embedding. In fact, Generative AI is not a value-neutral "glass jar" but "colored glasses" embedded with developers' values, training data biases, and the logic of capital pursuit. Without necessary scrutiny and regulation, the "micro-power" of algorithms is highly likely to dissolve the "macro-values" of ideological and political education, leading to "technological overreach" and "value loss." Therefore, how to maintain clear ethical consciousness while embracing technological dividends, complete the "disenchantment" of algorithms, and achieve the "reshaping" of educational values has become a major theoretical and practical issue that needs to be urgently responded to in the reform of university ideological and political education in the new era.

2. ALGORITHMIC DISENCHANTMENT: AN ONTOLOGICAL SCRUTINY OF EMBEDDING GENERATIVE AI INTO CIVICS EDUCATION

The term "Disenchantment" originates from Max Weber, meaning to strip away the mysterious veil of things and restore their rational essence. In the field of ideological and political education, "disenchanted" Generative AI means breaking the blind worship and deification of technology and deeply recognizing its essential limitations as a "probabilistic model" and the tension between it and the attributes of Civics education.

2.1. From "Omniscience" to "Probabilistic Statistics": The Dimensional Reduction of Cognition

In the cognition of many students, Generative AI is viewed as an omniscient "sage" or the "incarnation of truth." However, from the perspective of technological ontology, the essence of Generative AI (such as Large Language Models, LLMs) is a "probabilistic statistical prediction machine" based on massive corpora. It does not truly "understand" the meaning of the text it generates but predicts the next most likely word based on the preceding context (Next Token Prediction).

The core of ideological and political education lies in the faith in and pursuit of Marxist truth, which possesses rigorous logical necessity and historical inevitability. However, the content generated by AI is often a "platter" based on correlation rather than a "deduction" based on causality. When students become accustomed to replacing the deep exploration of "truthful knowledge" with AI's "probabilistic answers," the scientific nature and seriousness of ideological and political education face the risk of dissolution. We need to recognize that AI can provide "information" but cannot produce "wisdom"; it can simulate "logic" but cannot possess "faith."

2.2. From "Value Neutrality" to "Data Bias": Implicit Colonization

Technological neutrality theory argues that technology is merely a means, neither good nor evil. However, the training data for Generative AI is mostly derived from public text on the Internet. Given the current distribution structure of global internet information, English corpora dominate and are filled with values of Western-centrism, neoliberalism, and consumerism.

When data embedded with Western ideological biases is used to train models and applied to Civics classrooms in Chinese universities, algorithms become implicit "ideological colonizers." For example, when explaining core political concepts such as "Democracy," "Freedom," and "Human Rights," general large models without localized fine-tuning often output content conforming to Western political science definitions while excluding or distorting interpretations of Socialist Core Values [2]. This "algorithmic bias" is highly concealed; cloaked in "scientific objectivity," it subtly reshapes the value cognition of college students and poses a potential threat to national ideological security.

2.3. From "Subject Interaction" to "Simulacrum Interaction": The Disguise of Emotion

Civics Courses are education that "touches the soul," emphasizing sincere emotional communication and personality inspiration between teachers and students. Although Generative AI can simulate tones of "care," "encouragement," and even "humor" through affective computing technology to conduct fluent dialogues with students, this emotion is essentially an unconscious "Simulacrum."

Baudrillard pointed out that simulacra not only imitate reality but even replace it. In human-machine synergistic teaching scenarios, students may indulge in interactions with AI virtual characters, obtaining a cheap, instant emotional satisfaction, thereby escaping complex interpersonal interactions and profound emotional experiences in reality. This "painless" simulacrum interaction not only fails to replace the educational process of teachers "influencing life with life" but may also lead to the degeneration of students' emotional capabilities and alienation from real social interactions, causing "emotional hollowing" in ideological and political education.

3. RISK MAP: ETHICAL DILEMMAS IN THE APPLICATION OF GENERATIVE AI

In specific teaching practices, the ethical risks of Generative AI do not exist abstractly but materialize into realistic dilemmas in three dimensions: epistemology, axiology, and subjectivity.

3.1. Epistemological Dimension: Superposition of Information Cocoons and "Machine Hallucination"

(1) Historical Nihilism Triggered by Machine Hallucination:

Since the generation mechanism of Generative AI is based on probability rather than fact-checking, it often produces the phenomenon of "solemnly talking nonsense," known as "Machine Hallucination." In Civics teaching, this may manifest as fabricating non-existent historical events, inventing famous quotes, or even distorting Party history and national history. For college students lacking historical knowledge reserves and discernment, this seemingly detailed and logically self-consistent false information is highly deceptive, easily leading to the breeding of historical nihilism and shaking their historical confidence.

(2) Information Cocoon Effect Caused by Personalized Recommendation:

Generative AI excels at generating personalized content based on users' historical interaction data. While this improves the pertinence of teaching, it may also exacerbate the "Information Cocoon" effect. If a student reveals a certain erroneous value tendency (such as extreme individualism) in dialogue with AI, the AI, in order to cater to the user for positive feedback, may continuously generate content reinforcing this tendency, thereby confining the student in narrow cognitive biases and hindering the formation of a comprehensive and objective worldview [3].

3.2. Axiological Dimension: The Game between Algorithmic Black Boxes and Mainstream Ideology

(1) Loss of Interpretative Power of Values Due to Algorithmic Black Boxes:

The "black box" characteristic of deep learning models makes it difficult for us to trace the logical path by which AI generates a specific viewpoint. In the Civics classroom, when AI provides an explanation for a social hot issue, it is difficult for teachers to judge whether this explanation is based on a Marxist standpoint or some unknown algorithmic weight. If AI is allowed to control the "interpretative power" of complex social issues, the dominance of Civics Courses faces the risk of being sidelined, and the dissemination of mainstream ideology will be subject to uncontrollable algorithmic logic.

(2) Overreach of Instrumental Rationality over Value Rationality:

The core logic of Artificial Intelligence is "Instrumental Rationality" pursuing optimal efficiency. In AI-assisted teaching evaluation, there is often a tendency to quantify students' ideological and moral qualities into data indicators (such as interaction frequency, assignment duplication check rates), while ignoring substantive examinations of their inner world, moral emotions, and political beliefs. This "data-only" evaluation orientation easily guides students to become utilitarian and opportunistic, causing ideological and political education to deviate from the original intention of "fostering character and civic virtue" and falling into the trap of "calculative education."

3.3. Subjectivity Dimension: Dual Alienation of Teacher and Student Roles

(1) "De-skilling" of Teachers and Loss of Subjectivity:

With the comprehensive penetration of AI in lesson preparation, teaching, and Q&A, some teachers may develop severe technological dependence, becoming "operators" clicking mice or "porters" of AI content. Teachers may no longer delve into classic original works or painstakingly design teaching logic, but directly copy lesson plans generated by AI. This "de-skilling" not only weakens the professional authority of teachers but also leads to the loss of their educational subjectivity, causing teachers to degenerate from "engineers of human souls" to "technological vassals" [4].

(2) "Outsourcing of Thinking" and Lack of Criticality in Students:

The convenience of Generative AI tempts students to "outsource" the thinking process to algorithms. Facing complex theoretical problems or social phenomena, students no longer engage in independent deep thinking and logical deduction but habitually seek standard answers from AI. Over time, students' critical thinking abilities, innovative capabilities, and ability to solve complex problems will gradually atrophy, eventually becoming "digital giant infants" fed by algorithms, losing the subjective responsibility that newcomers of the era should possess.

4. VALUE RESHAPING: GOVERNANCE MECHANISMS FOR EMBEDDING GENERATIVE AI INTO CIVICS EDUCATION

Facing the ethical challenges brought by Generative AI, we must neither give up eating for fear of choking by simply shutting it out, nor let it drift by allowing technological logic to dominate education. We must adhere to the fundamental principle that "technology serves education," build a multi-dimensional governance mechanism, achieve a return from "technology-led" to "value-led," and complete the value reshaping of ideological and political education.

4.1. Technological Governance: Constructing an Independent and Controllable "Red Algorithm" System

Technological problems first need to find solutions from the technological source, achieving value alignment of algorithms through "controlling technology with technology."

(1) Building a High-Quality Corpus Exclusive to Ideological and Political Education:

Data is the "fuel" of AI. Universities should unite with research institutions and mainstream media to build an exclusive corpus covering Marxist classics, important documents of Xi Jinping Thought on Socialism with Chinese Characteristics for a New Era, materials on Party history and national history, and excellent Civics teaching cases. Strict cleaning, labeling, and classification of the corpus should be conducted to ensure the political nature, scientific nature, and purity of the training data, purifying AI's "cognitive genes" from the source.

(2) Developing Vertical Models Based on Value Alignment:

Relying on domestic independent and controllable basic large models, develop vertical domain Generative AI models for university ideological and political education. In the pre-training and fine-tuning stages of the model, introduce the "Reinforcement Learning from Human Feedback" (RLHF) mechanism. Organize Civics experts and ethics experts to form a "labeling team" to judge and correct AI-generated answers, embedding the rules of Socialist Core Values into the reward function of the algorithm, making AI learn to "speak politics" and "understand rules," ensuring its output content remains highly consistent with mainstream ideology [5].

(3) Establishing Full-Process Algorithmic Supervision and Circuit Breaker Mechanisms:

Develop an "Algorithmic Supervision System" tailored for Civics teaching scenarios. At the input end, set up sensitive word filtering and interception of inductive questions; at the output end, establish a real-time content review mechanism to scan AI-generated text and images for political security. Once high-risk content involving historical nihilism or ideological infiltration is found, immediately trigger the circuit breaker mechanism and conduct source tracing and accountability to ensure the absolute security of the teaching position.

4.2. Subject Governance: Reshaping Digital Literacy and Subject Consciousness of Teachers and Students

Humans are the masters of technology, and the core of governance lies in enhancing human subjective power and building a new human-machine subject relationship.

(1) Empowering Teachers: Transforming from "Knowledge Transmitters" to "Value Gatekeepers" and "Wisdom Guides."

Universities should launch an "AI Literacy Enhancement Project" for Civics teachers. It is necessary not only to train teachers to master the operational skills of AI tools but also to improve their "algorithmic awareness" and "ethical judgment." Teachers should become "gatekeepers" of AI-generated content, possessing the ability to keenly identify "machine hallucinations" and "algorithmic biases"; at the same time, they should become "guides" for student-AI interaction, adept at using erroneous cases generated by AI as negative teaching materials to conduct critical teaching, transforming "human-machine conflict" into "educational opportunities."

(2) Guiding Students: Growing from "Passive Recipients" to "Critical Thinkers" and "Human-Machine Collaborators."

In Civics teaching, special topics on "Digital Ethics" and "Media Literacy" should be added. Educate students to correctly recognize the instrumental attributes and limitations of AI and break superstitions about technology. Encourage students to always maintain a skeptical and scrutinizing attitude when using AI for learning, adhering to "human brain as primary, computer as auxiliary."

Through teaching activities such as "Human-Machine Debate" and "Finding Faults in Algorithms," exercise students' critical thinking and cultivate their ability to make independent value judgments and moral choices in the intelligent era [6].

4.3. Institutional Governance: Perfecting Ethical Regulations and Evaluation Systems for Human-Machine Synergy

Institutions are the guarantee of governance, drawing red lines and bottom lines for technology application through establishing rules and regulations.

(1) Issuing "Guidelines for the Ethical Application of Generative AI in University Civics Courses":

Led by educational authorities, formulate national guidance documents clarifying the application principles, scope, and prohibited behaviors of Generative AI in Civics teaching. For example, explicitly stipulate that issues involving major political principles and historical verdicts must rely on central documents and unified textbooks, and direct use of AI-generated content is prohibited; clarify the limits on the proportion of AI assistance in student assignments and labeling obligations, and severely punish academic misconduct.

(2) Constructing a "Dual-Teacher Synergistic" Teaching Responsibility System:

Clarify the auxiliary status of AI as a "Digital Teaching Assistant" and establish the responsibility principle of "Human Teacher as the Bottom Line." Regardless of how advanced technology becomes, human teachers are always the primary persons responsible for teaching quality and ideological security. Establish a filing system for human-machine synergistic teaching, keeping traces and archives of AI-generated lesson plans and courseware for ethical review and teaching assessment.

(3) Reconstructing a Multi-Dimensional Evaluation System Embodying "Value Rationality":

Reform the quantitative evaluation mode that relies solely on data and build an evaluation system that "emphasizes both process and result, and integrates data with values." While using AI for learning situation analysis, increase the weight of qualitative evaluations such as teacher-student interviews, behavioral observations, and open writing. Focus on examining students' political identification, depth of moral emotion, and ability to translate theory into practice, ensuring that the evaluation system always serves the fundamental goal of "fostering character and civic virtue" and prevents being hijacked by instrumental rationality.

5. CONCLUSION: TOWARDS A NEW EDUCATIONAL ECOSYSTEM OF "HUMAN-MACHINE CO-GOODNESS"

Heidegger once said: "But where danger is, grows the saving power also." [7] The intervention of Generative AI brings unprecedented ethical risks to university ideological and political education, but it also forces us to re-examine the essence of education and ponder the ultimate propositions of "what makes a human" and "what is education."

The essence of ideological and political education is the awakening of the soul and the transmission of values. This is a spiritual practice full of warmth, wisdom, and conviction, which no algorithm can completely replace. The purpose of algorithmic disenchantment is not to deny technology but to return technology to its position as a "tool"; the aim of value reshaping is not to reject intelligence but to make intelligence serve the development of "humans."

Looking to the future, university ideological and political education should remain clear-headed in "algorithmic disenchantment" and firm in confidence in "value reshaping." Through the synergistic governance of technology, subjects, and institutions, we can build a new educational ecosystem of "Human-Machine Co-Goodness." In this ecosystem, AI provides the breadth and efficiency of knowledge, while human teachers provide the depth and warmth of thought; technology is no longer

an alienating force but a ladder leading to truth. We have reason to believe that Generative AI, after ethical regulation and value guidance, will eventually become the "spark of wisdom" that lights up students' lamps of faith, promoting university Civics Courses to radiate more brilliant vitality in the intelligent era.

ACKNOWLEDGEMENTS

Research Outcome of the 2025 Higher-Education Study Project under the 14th Five-Year Plan of the Guangdong Higher-Education Association (Project No. 25GQN011): “Mechanisms and Pathways for Constructing Scenario-Based Teaching Models for AI-Enabled Ideological and Political Education in Universities Empowered by Generative AI”

REFERENCES

- [1] Marx, K. (2004). *Das Kapital* (Volume I). Beijing: People's Publishing House, p. 487.
- [2] Feng, G., & Sun, W. (2023). Generative Artificial Intelligence Empowering Ideological and Political Education: Internal Mechanism, Risk Challenges, and Practical Paths. *Studies in Ideological Education*, (12), 112 -118.
- [3] Luo, Y., & Guo, X. (2022). Scenario-based Ideological and Political Education: Connotation, Value, and Path. *Dao Guide of Ideological and Theoretical Education*, (09), 108-114.
- [4] Wu, M., & Dong, J. (2020). Challenges and Responses to Ideological and Political Education in the Algorithmic Age. *Studies on Marxism*, (05), 134-143.
- [5] Zhang, Y. (2021). *Intelligent Civics: Research on Innovation of Ideological and Political Education in Universities in the Big Data Era*. Beijing: People's Publishing House, pp. 45-50.
- [6] Lan, G., Guo, Q., Wei, J., et al. (2023). Generative Artificial Intelligence Empowering Digital Transformation in Education: Opportunities, Challenges, and Paths. *Open Education Research*, 29(03), 25 -33.
- [7] Heidegger, M. (2015). *The Question Concerning Technology* (Z. Sun, Trans.). Shanghai: Shanghai Joint Publishing Company, p. 35.