

The Appropriateness and Boundary Governance of Embedding Artificial Intelligence Technology into College Students' Value Guidance

Jing Ren¹, Xiaoguang Sun^{2*}

¹Inner Mongolia Technical College of Mechanics and Electrics, Hohhot, 010070, China

²Inner Mongolia University of Finance and Economics, Hohhot, 010070, China

*Corresponding author: sunxg@imufe.edu.cn

Abstract: With the spread of artificial intelligence technology in the areas of information dissemination and cognitive interaction, it has also begun to be applied in guiding the core socialist values of college students, and now it is a new force influencing the value construction of young people. Technology embedding is not without flaws; although it can enhance the efficiency of information dissemination, there are also problems, such as the digital alienation of the guidance subject, harm to value autonomy, and the substitution of value rationality with instrumental rationality. According to the view of Marxist theory, this paper will examine the logic of appropriateness by studying the theoretical threshold of compatibility between technological functions and human subjectivity, the structural coupling of mediating effects and value goals, and the free and all-encompassing development of people. There is a contradiction at the boundary of the substitution of guiding subjectivity, algorithmic cognitive discipline and conflict with value autonomy. Some governance ideas have been put forward, such as context-specific hierarchical regulation, explainable constraints based on core socialist values, division of labour and feedback control in human-machine collaboration. Establish a theoretical limit for the moderate integration of artificial intelligence in the guidance of core socialist values and thus achieve the goal of cultivating a new generation with the ability to carry out the mission of national rejuvenation.

Keywords: Embedding of Artificial Intelligence Technology, Value Orientation, Appropriateness, Boundary Tension, Governance Mechanism

Introduction

The goal of guiding the values of college students is to help them consciously learn value norms and consistently apply them in life; thus, one should take into account all aspects of a person's cognitive state, emotions and circumstances. AI technology has been used to develop some new technologies that can help organise information, predict people's needs and understand language in the area of guidance. There is an endogeneity in the computable logic of technological systems that is not in line with their nonlinear and contextual nature in value guidance. If there is too much dependence on technology, we will have to replace expert judgment in the area of guidance with algorithm recommendations; thus, the principle of individual autonomy will be violated and quantitative translation of guidance will not be achieved. Based on the above, we believe that it is necessary to study the logic of appropriateness and boundary governance for embedding artificial intelligence in value guidance, clarify the conditions and limitations of technological adaptation, and at the same time build a governance framework to address the risks of such embedding. The three dimensions of this study are appropriateness judgment, identification of boundary tension and design of a governance mechanism; together they will help us understand how to integrate technology with the basic logic of guidance.

1. The Logic of Appropriateness for Embedding Artificial Intelligence Technology in College Students' Value Education Guidance

1.1 Application of Artificial Intelligence Technology in the Area of Guidance Function

The first way artificial intelligence technology can help spread the core socialist values is to organise the architecture of information flow and increase the frequency of cognitive exposure.

Collaborative filtering and content tag matching are used to generate all kinds of information sequences based on people's interaction patterns, and thus the mode of transmission for value symbols is changed from unilateral broadcast to multi-node distribution; programming of this technology can be employed to present multimodal content with predetermined weights and achieve dynamic coordination among guiding goals and cognitive inertia in the dissemination of information. Adaptation can increase efficiency, and at the group level, it can also alter the diffusion topology of value information to enable differentiated allocation of guidance resources according to the different cognitive states of individuals, and it can be shown that technology can extend the essential powers of humanity^[1].

The degree of functional adaptation is that Natural Language Processing and Affective Computing can be used to achieve semantic understanding and emotion recognition. Quantifying the trend of attitude and emotional poles in people's speech can help us find potential deviations in their values under non-interventional conditions; for example, we can identify signs of alienation from patriotism, dedication, integrity and friendliness. This identification should not be taken as the expert opinion of the guidance practitioner; rather, it is to be used as a cognitive aid to reduce the blind spots in guidance caused by information asymmetry. Therefore, technological adaptability refers to all kinds of mappings of information flow, emotional flow and value flow, and its basic criterion is whether it can maintain and improve the subject position of people; thus, technology should genuinely serve to meet the high demands of all-round development for people.

1.2 The Coupling of Value Guidance Objectives and Technological Mediation Effects

The goal of leading in core socialist values is to help college students internalise the goals of their own lives and learn about the requirements of society and daily life. Artificial intelligence is an intermediary, so the effect of technological mediation can be observed in how information is filtered, organised and presented; the parameters of the algorithm directly determine the gradient of visibility for value-related information. It is not mediation by neutral technology; rather, it is an algorithmic distribution of weights for the purpose of guidance. The first necessary condition for preventing instrumental rationality from replacing value rationality is that the algorithmic parameters can be guided by the demands of values.

The construction of the coupling relationship should meet the condition that there is a structurally mappable similarity between the parameter setting of technological mediation and the value orientation of guidance objectives. That is to say, the value hierarchy of core socialist values needs to be expressed in a quantifiable way in the loss functions or reward mechanisms of algorithms, but at the same time, it should not lose its rich connotations due to quantification. For example, the diversity index of the value information can be added as a regularisation term in a recommendation system to avoid the problem of excessive accumulation of homogeneous content and ensure that all kinds of values, such as justice and harmony, are presented in an all-round way. When the technology has achieved stable production of semantic bias in line with certain values, such as patriotism, dedication and integrity, then functional coupling has been realised. This link does not mean that all choices of individual values have been ruled out; rather, it can help the technological system leave some space for the continuous spread of core socialist values in its parameter space and allow people to make free choices within the reasonable information boundaries set by the system, thus accumulating subjective strength for progress in free and open development.

1.3 Multidimensional Theoretical Basis and Threshold Delineation for Appropriateness Judgment

The foundation of the judgment of appropriateness should be the Marxist critique of technological alienation, the philosophy of human subjectivity and the theory of free and all-round development of people. In terms of the philosophy of technology, as a tool for cognitive extension, artificial intelligence should be used to enhance the interpretative authority of guidance practitioners and not to replace them; from the perspective of the philosophy of subjectivity, technological embedding should not damage metacognitive monitoring or reduce people's freedom of choice; and in light of the theory of all-round development, technology should not narrow the process of value guidance or weaken emotional attachment. The three factors form a theoretical triangle, and the purpose is to determine whether they can promote the development of new-generation people with independent personalities and good values.

Based on the above theoretical basis, we can set three levels of suitable thresholds. The first threshold is whether the technological intervention causes a semantic distortion of the guidance objective; that is to say, whether the interpretability of the value information after algorithmic

processing is too low and results in a misinterpretation of the core connotations of socialist values. The second is to determine how much the technological system has reduced the space for professional judgment in guidance work; if the force of algorithmic recommendations is stronger than the veto power of guidance practitioners, it may be considered that the technology has shown signs of transgression and alienation. The third threshold is the extent of harm to people's value autonomy; if people cannot explain the reasons for their own choices of values in retrospect, it means that technological embedding has gone too far and begun to deviate from the goal of free and all-round development for all people. The threshold division is not a fixed number but a system of boundaries that changes according to the development of technology and changes in the environment for guidance; its calibration frequency should be in line with how the system operates, and it needs to be continuously adjusted to protect the interests of the people^[2-5].

2. Boundary Tensions of Embedding Artificial Intelligence Technology in College Students' Value Education Guidance

2.1 Possible Ways Technology May Affect the Subjectivity of Guidance

The first appearance of the erosion of guiding subjectivity due to technological invasion is the substitution of subject functions in the process of decision-making delegation. With the help of artificial intelligence, we can filter information, predict people's needs, and plan lives in accordance with values, agency and context; thus, the demand for human guidance is declining, and only uniform algorithmic recommendations are needed. It is not the free handoff of power, but rather the accumulation of path dependence due to daily guidance operations that has reduced decision-making costs. Guidance practitioners will be more likely to use the content strategies and forms of interaction provided by the system; they will reduce their own differentiated judgments based on professional experience and relational sensitivity, and thus the subjectivity of guidance will be replaced by the presupposed logic of the technological system.

Another reason for the reduction in supervision is that the technology is a "black box" and therefore not easy to understand. Given that the parameters and many non-linear transformations in AI models are large, it is not easy to trace the cause-and-effect path from input to output. Guidance practitioners do not know why the system recommended a particular order for the values in the information, so they cannot carry out appropriate supervision and guidance. As they cannot determine whether the technological output is in line with the guiding goals, the regulatory power of the guidance practitioner is lost, and subjectivity gradually gives way to the autonomous operation of the technological system. It did not happen all at once; rather, it is the result of many structural deviations caused by the repeated use of technology.

2.2 Analysis of the Conflict between Algorithmic Cognitive Discipline and Value Autonomy

The first basic operation mode of algorithmic cognitive discipline is to adjust how people distribute their attention and how well they remember particular values through repeated presentation and feedback. The information flow generated by the behaviour of people in recommendation systems is constantly in line with the same direction of values, and it can serve as an adjustable cognitive guide. As a result of the above mechanism, the range of choices for people has been narrowed, and many of the perspectives required for value judgement have also gone missing. At the base of value autonomy is one's ability to compare, weigh and make conscious choices among many options in life; however, algorithmic discipline reduces the number of comparison choices by limiting them and thus creates a structural conflict with value autonomy^[6].

Another way to show this conflict is that people do not know where their own values come from. Those who have lived in an algorithmic world for a long time often do not know whether their sense of values comes from genuine independent thought or from passive assimilation of repeated system recommendations. When people believe that the cognitive inertia of the technology system is an extension of their own will, they have lost value autonomy without being aware of it. Therefore, one cannot simply turn off the algorithmic recommendation; it has been deeply embedded in how people obtain and judge information. In short, the aim of the system (to promote user participation or ensure the accuracy of information) does not cover all the different demands for openness in the spirit of autonomy.

2.3 Manifestations of Boundary Disorder Induced by the Computability of the Guidance Process

The computability of the guidance process first appears as errors in the quantitative translation of value objectives. The scope of emotional identification, the construction of meaning and the internalisation of value guidance are not independent or linear; thus, they cannot be fully represented by computable index variables. When the AI system breaks down the guiding objective into proxy indicators, such as click-through rate, dwell time and sentiment polarity score, there will be a semantic gap between the measured targets and the actual goals. Translation errors will cause fluctuations in the measurable proxy variables, but they will not affect the actual change in the effect of the guidance provided by the technology system; therefore, a boundary disorder will occur and instrumental rationality will be separated from value rationality in the guidance process.

The second kind of boundary disorder is a clash between the independent operation of the technology system and the actual circumstances of the guidance situation. Value guidance is based on an all-encompassing understanding of people's current situations, relationships, surrounding environments, etc., but artificial intelligence systems can only process organised features that have been pre-processed. If only a few attributes are used by the technology system to produce the guidance recommendations, the decision-making logic will be too general for the specific circumstances. If the guidance practitioners do not follow the recommendations of the system, the results will not meet the original goals of the guidance; otherwise, the purpose of technological intervention will not be achieved. This problem shows that the boundary of the guidance process has moved from functional assistance to functional obstruction, and it is no longer possible to integrate the operating logic of the system with the internal logic of guidance^[7].

3. Boundary Governance Mechanisms for Risks of Inappropriate Embedding

3.1 Hierarchical Regulation of Technological Intervention According to Guidance Contexts

At different times, the degree of technological support will have different impacts on the functions of the system, and the goal is to safeguard the ideological security of the country and the individuality of the people. Based on the sensitivity of the information, the required degree of autonomy and the depth of emotional involvement, different contexts exist; for example, there is dissemination of low-sensitivity information and attitude negotiation, and value conflicts are another kind. The lowest level is to collect and organise information; at the level of attitude negotiation, several options can be offered, but the final choice of the guidance worker must still be made; finally, at the highest level of resolving value conflicts, only help in recognising emotions is provided, and prescriptive suggestions based on values are not allowed. The hierarchy of division is based on how much the guiding goal depends on the subjective judgment of teachers; the higher this dependence is, the lower the authority of technological intervention should be, and thus we need to avoid excessive instrumental rationality.

Develop an Operating Index System for Context Classification in the Implementation of Hierarchical Regulation based on the Experience of Ideological Work at Higher Education Institutions. The three sections of the index are: value cognitive stability, degree of value consensus and density of emotional expression. It is particularly necessary that, for ideological issues related to important matters of principle, the algorithm does not independently generate value-judgment content; it should only spread authoritative information that has been strictly manually verified, and the system should be reduced to an information-transmission channel. If the risk of the task is high, then only the passive monitoring mode of the system will be employed. Hierarchical regulation is not uniform across all types of technological functions; instead, under different circumstances, various levels of authorisation are given according to the degree of ideological risk to ensure that the technology serves the main purpose of moral education and talent cultivation.

3.2 Value-oriented Constraint Principle of Technological Explainability

In accordance with the constraint principle of technological explainability, when a system is employed to guide the implementation of core socialist values, the foundation for decision-making and the rationale behind the output must be intelligible and traceable for the guidance practitioners, and this information should be presented in the framework of the discourse of core socialist values. It should not be confused with the general problem of algorithm explainability; rather, it should show how the output information is causally linked to the desired value goal in an understandable way, not just the

parameters. That is to say, when the system recommends content, it needs to output both the judgment path of value relevance and an explanation that is in line with the value requirements of patriotism, dedication and integrity; it should not only provide a list of feature weights. Therefore, the guide can determine whether the values of the system are in line with the main socialist ideas.

The particular construction of the constraint principle can be achieved by means of the two technological paths of attention mechanism visualisation and counterfactual explanation. Visualisation of the attention mechanism can show which semantic units are given more weight by the model in processing the input information and reveal the basis for filtering value-related information, and counterfactual explanations can show the necessary conditional relationship among value elements and system decisions through comparison. Both kinds of explanations should be in simple language, and the basic vocabulary should include socialist values; for example, do not use technical charts or mathematical symbols, but rather say that "this content is recommended because it embodies the values of justice and friendliness, and certain content is suppressed because it poses a risk of contradicting integrity." The purpose of the explainability constraint is to help the guidance practitioner reasonably judge the value-judgment logic of the system with little mental effort and promptly correct any errors that have occurred. Any technological system that does not meet the above basic requirements should not be used to guide values in practice, even if it has a good algorithm performance, in order to prevent the logic of technology from overriding human values^[8-10].

3.3 Division of Responsibilities and Feedback Calibration in Human-Machine Collaborative Guidance

The Division of Responsibilities should consider the characteristics of higher education institutions and divide the tasks of algorithm design, system deployment and guidance implementation according to the differences among them. The development party needs to design the algorithm and, in doing so, should be aware of the values in society and avoid bias. As the introducing organisation will be responsible for operating the system, it should consider the actual circumstances and set up a regulation hierarchy. Finally, the leading body of the institution is responsible for the general leadership of ideological work in this area, and frontline ideological and political educators should carry out guidance, conduct prudent reviews and make final decisions on the results of technology. The three parties are linked; if any link is missing, we will not be responsible, and there will be a system failure in combating digital alienation.

The purpose of the feedback calibration mechanism is to address the problem of behavioural drift caused by changes in the technological system over time, and it is a type of collaborative governance involving many parties. As the distribution of data and models changes or is updated, the original hierarchical regulations and explainability constraints will gradually lose their effectiveness; therefore, it is necessary to build a bidirectional regulation loop: the usage experiences and judgment results of guidance practitioners will be regularly collected in an organised way as feedback and used to adjust the constraint parameter module, and at the same time, changes in system behaviour will be continuously monitored through monitoring indicators, and manual review will be triggered when necessary. A joint mechanism should be established to carry out regular joint assessments of the ideological trends in the class by teachers of the ideological and political course, counselors and student leaders, incorporate the practical wisdom of frontline educators and authentic feedback from students into the calibration loop, and thus prevent the superposition and amplification of individual cognitive biases and deviations in the technology system through collective assessment. The next calibration will be carried out after a certain period of time has passed since the last one. At a reasonable level, human-computer collaboration can be achieved by dividing the division of responsibilities and allowing for flexible cooperation among all parties; artificial intelligence should be employed as an auxiliary tool to support people in all aspects of their development and not operate independently.

Conclusion

Based on Marxist theory, this paper has constructed an analytical framework for the integration of artificial intelligence in the education of core socialist values for college students, and the three levels of this framework are the logic of appropriateness, boundary tensions and governance mechanisms. In terms of appropriateness, it has been shown that the functional adaptation of technology should be based on the improvement of human subjectivity, and its goal should be the free and all-round development of all people; at the level of boundary tensions, problems such as digital alienation,

algorithmic discipline and the usurpation of value rationality by instrumental rationality have been pointed out, and their origin can be traced back to the incommensurability between the objective function of technological systems and the logic of comprehensive human development; at the level of governance mechanisms, three ways have been proposed: context-sensitive hierarchical regulation, explainability constraints based on the discourse of core values, and division of responsibilities with feedback calibration. The deeper inquiry of this study is what kind of technological application can truly serve the cultivation of the new generation and carry the spirit of undertaking the mission of national rejuvenation. In the future, research will focus on the adaptive calibration of appropriateness thresholds, differentiate the response studies of all kinds of value dimensions, build responsibility attribution models based on Marxist subjectivity theory, carry out comparative analyses of alienation risks under different technological architectures, and make artificial intelligence a true supportive force for the all-round development of people.

Fund Projects

The research project titled "Research on the Precision Teaching Model of Human-Machine Collaboration for the 'Corporate Culture' Course in the Context of Digital Transformation" under the Educational Science Planning Project of the Inner Mongolia Autonomous Region has been concluded, with project number: NGJGH2024195.

References

- [1] Huang Xinrong, and Li Run. "Experimental Ideological and Political Education: Improvement and Innovation of the Disciplinary System of Ideological and Political Education in the Age of Digital Intelligence." *Ideological and Theoretical Education* 2026(3): 52-58.
- [2] Shi Yunxia, and Zhou Jinxin. "A Study on the Innovation of Ideological and Theoretical Education of the Communist Party of China in the New Era." *Guide to Ideological and Theoretical Education* 2026(3): 134-143.
- [3] Feng Gang, and Yu Yifei. "The Disciplinary Significance of Ideological and Political Work: Theoretical Elucidation, Empirical Reflections, and Development Prospects." *Ideological and Political Education Research* 42(1): 1-7.
- [4] Ai Jun. "Discourse Innovation and Path Optimization of Ideological and Political Education in Colleges and Universities under the Background of Youth Subculture." *Ideological Education Research* 2026(2): 138-143.
- [5] Xu Rong. "Some Reflections on the Construction of the Autonomous Knowledge System of the Discipline of Ideological and Political Education in the Age of Digital Intelligence." *Research on Marxist Theory Disciplines* 11(10): 94-103.
- [6] Dai Yuqi. "The Leading Force of Ideological and Political Education: A Signature Concept of the Discipline of Ideological and Political Education." *Research on Marxist Theory Disciplines* 11(8): 105-112.
- [7] Zeng Linghui. "On the Innovative Development of the Basic Theories of the Discipline of Ideological and Political Education in the New Era." *Ideological and Theoretical Education* 2025(3): 49-56.
- [8] Zhou Ke. "The Contemporary Interpretation of Marx's Theory of Alienation and Its Significance." *Theoretical Exploration* 2025(1): 13-21.
- [9] Shen Guangming. "The Ecological Narrative of Marxist Theory of Productive Forces." *Marxist Studies* 2025(11): 67-74.
- [10] Zeng Ruiming. "The Scientific Clarification of the Concept of Marxist Social Revolution and Its Theoretical Effects." *Marxist Studies* 2025(1): 41-51.