Generative AI-supported Deep Integration Assessment and Support Strategies for Curriculum Ideological and Political Education

Qinlin Cai*, Youyou Xiang, Feiyu Zhan
Zhejiang Normal University, Jinhua 321004, China
*Corresponding email: 2897245759@qq.com

Abstract

Curriculum ideological and political education (CIPE) has emerged as a critical approach to fostering moral character and ideological awareness in higher education, yet its implementation often suffers from superficial integration and fragmented practices. This study proposes a generative artificial intelligence (AI)-supported assessment framework for evaluating the deep integration of CIPE across disciplinary courses and develops targeted support strategies to enhance integration effectiveness. Drawing upon cognitive transfer theory, value internalization models, and the Technological Pedagogical Content Knowledge (TPACK) knowledge framework, this study first constructs a multi-dimensional evaluation indicator system encompassing four core dimensions: political guidance, cultural identity, ethical norms, and practical innovation. Employing the Analytical Hierarchy Process combined with the Delphi method, this study establishes a three-level indicator structure equipped with adaptive quantification algorithms to accommodate disciplinary differences. Subsequently, this study develops a multimodal assessment methodology leveraging large language models through prompt engineering, implementing a dual-cycle assessment system enhanced by proximal policy optimization reinforcement learning. The framework integrates Bidirectional Encoder Representations from Transformers and Long Short-Term Memory (BERT-LSTM) hybrid neural networks to generate self-explanatory integration index cloud maps. Furthermore, this study designs an intelligent support strategy framework that dynamically generates personalized prompts and triggers adaptive intervention mechanisms. A quasi-experimental study conducted across six university courses demonstrates that the proposed assessment method achieves 87.3% consistency with expert ratings (Cohen's κ=0.82), and the intervention strategies yield a 23.7% improvement in CIPE integration depth compared to control groups. This research contributes both theoretically and practically to the intelligent evaluation and targeted enhancement of curriculum's ideological and political integration in the digital education era.

Keywords

Curriculum ideological and political education, Generative artificial intelligence, Deep integration evaluation, Multimodal learning analytics, Support strategies

Introduction

The integration of ideological and political education into specialized curricula, commonly referred to as curriculum ideological and political education (CIPE), represents a fundamental strategic initiative in Chinese higher education aimed at cultivating morally grounded, socially responsible professionals. Since its formal introduction following the 2016 National Conference on Ideological and Political Work in Universities, CIPE has been widely implemented across disciplines, seeking to unify knowledge transmission with value guidance. However, despite substantial policy support and institutional investment over the past decade, the practical implementation of CIPE frequently encounters significant challenges, notably the prevalence of superficial hard integration approaches that merely append ideological content to existing curricula without achieving deep, organic fusion with disciplinary knowledge. This fragmentation not only diminishes educational effectiveness but also creates a disconnect between professional competency development and value formation, ultimately undermining the fundamental goal of fostering well-rounded graduates who possess both professional expertise and strong ideological character [1].

Contemporary assessment practices for CIPE remain largely reliant on subjective expert evaluation, questionnaire surveys, and static indicator checklists, which fail to capture the dynamic, multi-faceted nature of ideological integration in real teaching contexts. Traditional evaluation methods typically prioritize outcome measurement over process monitoring, offering limited diagnostic capacity and providing few actionable insights for instructional improvement. Moreover, the significant variation in how ideological elements manifest across different disciplines, from experimental ethics and scientific responsibility in engineering to cultural heritage and historical perspective in humanities, renders standardized evaluation instruments insufficiently adaptive. These limitations underscore the urgent need for more sophisticated, context-aware assessment approaches that can provide real-time, granular insights into CIPE integration quality while accommodating disciplinary differences in how ideological education manifests [2].

The rapid advancement of generative artificial intelligence has opened unprecedented possibilities for transforming educational assessment and instructional support. Large language models such as GPT-4, DeepSeek, and domain-specific variants have demonstrated remarkable capabilities in natural language understanding, content generation, and multi-modal analysis, offering new technical pathways for intelligent educational evaluation. In particular, prompt engineering techniques enable the design of sophisticated evaluation rubrics that can guide large language models (LLMs) to perform complex assessment tasks with human-like reasoning, while reinforcement learning mechanisms allow assessment models to iteratively improve through feedback. When combined with multimodal learning analytics that integrate text, audio, video, and behavioral data, generative artificial intelligence (AI) can potentially capture the richness of classroom interactions in ways that traditional methods cannot, offering both depth of analysis and scalability of application. These technological advances create a timely opportunity to address longstanding challenges in CIPE assessment through intelligent, data-driven approaches [3].

Building upon these technological advancements, this study addresses three critical research gaps in current CIPE research. First, existing evaluation indicator systems lack sufficient granularity and disciplinary adaptability, often relying on one-size-fits-all metrics that fail to account for domain-specific characteristics of ideological integration. Second, CIPE assessment remains predominantly manual and static, lacking real-time monitoring capabilities and intelligent diagnostic functions that could support ongoing instructional improvement. Third, there is a dearth of evidence-based support strategies that can translate assessment results into actionable instructional interventions, creating a persistent gap between evaluation and improvement action. Against this backdrop, the present study firstly aims to construct a fine-grained, discipline-adaptive evaluation indicator system for CIPE deep integration. Secondly, it intends to develop a generative AI-powered multimodal assessment methodology for real-time CIPE integration evaluation. Thirdly, it seeks to design and validate targeted support strategies for enhancing CIPE integration quality.

This research makes several significant contributions to the field. Theoretically, it integrates cognitive transfer theory, value internalization models, and educational ecology perspectives to establish a comprehensive conceptual framework for understanding deep CIPE integration. Methodologically, it introduces an innovative assessment pipeline that combines prompt engineering with reinforcement learning for educational evaluation. Practically, it provides a deployable intelligent assessment system and evidence-based intervention strategies that can directly support teachers and educational administrators in improving CIPE implementation quality. The following sections present the literature review, theoretical model development, research methodology, empirical results, and discussion of implications.

Literature review

Curriculum ideological and political education: concepts and evaluation

The conceptual foundations of curriculum ideological and political education draw from both Chinese educational philosophy and international perspectives on values education. Domestically, some researchers conceptualize CIPE as a curriculum perspective that emphasizes the permeation of ideological and political elements across all courses, forming a comprehensive ideological education structure. This view aligns with the holistic education paradigm articulated by Miller, which advocates for integrating intellectual, emotional, and ethical development across the curriculum rather than treating them as separate domains [4]. Du et al. further elaborated a three-in-one framework for university ideological education systems, aiming to break down the traditional silos between specialized ideological courses and professional education and create a more cohesive educational experience [5]. In terms of implementation models, researchers propose a three-stage process comprising element mining, instructional design, and evaluation feedback, which has become a dominant framework for CIPE practice in Chinese higher education institutions [6,7].

Disciplinary variations in implementation are well-documented in the growing CIPE literature. Science and engineering courses tend to emphasize scientific spirit and research ethics through error analysis cases and responsible innovation discussions, while medical programs integrate professional ethics centered on the value of life and patient-centered care. Humanities and social science courses, by contrast, often have more natural points of connection with ideological content through cultural analysis, historical perspective, and critical thinking frameworks [8]. Internationally, analogous concepts such as service learning, character education, and moral development theory provide cross-cultural references for understanding how values can be integrated into academic curricula, though these frameworks differ in their specific ideological orientations and educational objectives [9,10].

Evaluation of CIPE effectiveness has evolved from qualitative descriptions toward more structured quantitative approaches over recent years. Researchers have employed various methodologies including the Delphi method for expert consensus building, structural equation modeling for indicator construction, and adaptations of the Defining Issues Test for moral development assessment. Institutional implementations such as Zhengzhou University's Curriculum Ideological and Political Teaching Quality Evaluation Scale, developed within smart classroom environments, represent early attempts at technology-enhanced evaluation. However, as Leary and Sherlock observe in their bibliometric analysis of CIPE research, studies remain concentrated on indicator system construction with limited attention to assessment methodology innovation and practical tool development [11]. A critical limitation is the predominance of coarse-grained, static indicators that struggle to capture the dynamic, emergent nature of ideological integration in live classroom settings, along with insufficient disciplinary adaptation mechanisms that limit applicability across diverse academic fields.

Generative AI in educational assessment

The application of artificial intelligence in education has progressed considerably from early intelligent tutoring systems to contemporary generative AI-enhanced learning environments. Generative AI, characterized by its capacity to produce novel content that reflects patterns in training data, represents a paradigm shift in educational technology. Within the assessment domain, generative AI offers distinctive advantages over traditional educational data mining approaches, particularly in handling unstructured data, understanding contextual nuance, and generating human-readable evaluation explanations that support learning and improvement. Current research on generative AI in educational assessment encompasses several key directions: assessment ecosystem design, formative assessment pathways, and technical implementation architectures [12-15].

In assessment ecosystem design, Ren et al. propose a human-machine collaborative intelligent assessment ecology that supports multi-dimensional dynamic evaluation of learning processes, emphasizing the complementary strengths of human judgment and AI scalability [16]. In formative assessment pathways, some researchers construct human-centered formative evaluation paths drawing upon domain-specific LLMs such as Large Language-and-Vision Assistant (LLaVA), emphasizing multi-source triangulation of evidence for more robust assessment conclusions. In technical implementation, researchers have explored various architectures, including multi-modal data processing pipelines, neural collaborative filtering for group recommendation systems, and cross-modal explanation generation systems for language learning assistance.

Prompt engineering has emerged as a critical methodology for leveraging LLMs in educational assessment, with structured input prompts encoding evaluation criteria and domain knowledge guide models to perform complex assessment tasks [17,18].

The concept of dual-loop systems represents an advancement beyond static prompting, where assessment outputs are compared against expert ratings and used to refine both prompt templates and model parameters through techniques like Proximal Policy Optimization (PPO) reinforcement learning. This approach addresses a key limitation of early LLM applications, their tendency to produce inconsistent or biased evaluations, by incorporating iterative feedback mechanisms that progressively improve assessment quality. Despite these advances, several challenges remain in applying generative AI to educational assessment, particularly in values-laden domains like CIPE. The other researchers note that current assessment models still exhibit significant limitations in handling nuanced value judgments and contextual interpretation. The dynamic adaptability of generative AI assessment models for collaborative problem-solving competencies, the ethical boundaries of AI evaluation in sensitive ideological domains, and the optimization of human-machine collaborative mechanisms all require further investigation. Additionally, most existing research focuses on theoretical construction and technical verification, with limited longitudinal empirical studies in authentic classroom settings [19,20].

Multimodal analytics and support strategies

Multimodal learning analytics represents a growing field that integrates data from multiple sources, including text, speech, video, physiological signals, and interaction logs, to build comprehensive understandings of learning processes. For CIPE assessment, multimodal data is particularly valuable because ideological integration manifests not only in textual content but also in teacher discourse patterns, classroom emotional climate, student engagement behaviors, and nonverbal communication. The Bidirectional Encoder Representations from Transformers and Long Short-Term Memory (BERT-LSTM) hybrid neural architecture has shown promise for fusing heterogeneous data types, combining BERT's strength in semantic representation with LSTM's capacity for temporal sequence modeling. Data fusion techniques typically follow three approaches: early fusion combining raw features, late fusion integrating model outputs, and intermediate fusion merging representations at hidden layers. For CIPE assessment, intermediate fusion through multi-modal alignment networks offers particular advantages because it allows models to learn cross-modal correspondences, for example, linking specific linguistic patterns in teacher speech with corresponding classroom participation patterns [21,22].

Turning to support strategies for CIPE enhancement, research has primarily focused on instructional design guidelines, teacher training programs, and resource development. However, these approaches tend to be generic rather than personalized, providing uniform recommendations regardless of specific course contexts or identified weaknesses. The concept of adaptive intervention, dynamically adjusting support based on real-time assessment data, remains largely unexplored in the CIPE literature. In broader educational technology research, adaptive learning systems have demonstrated effectiveness through techniques knowledge tracing, dynamic difficulty adjustment, and personalized feedback generation [23,24]. The closed-loop improvement cycle draws from design-based research methodology, which emphasizes iterative refinement of interventions in authentic settings through cycles of design, implementation, evaluation, and revision. When combined with AI-powered diagnostic capabilities, this approach can create closed-loop improvement systems. Within such systems, assessment data directly informs targeted interventions, whose effects are measured and used to further refine assessment methods and intervention strategies. This study operationalizes this cycle for the CIPE context, developing support strategy categories including diagnostic prompts, affective motivation, and resource provision that can be dynamically triggered based on assessment outputs [25].

Model design and hypotheses development

Theoretical framework: Cognitive-affective-behavioral model

The theoretical foundation of this study rests on a three-stage educational objective model encompassing cognition, affection, and behavior, which provides a hierarchical structure for understanding how curriculum ideological and political integration achieves its educational goals. This framework draws from Bloom's taxonomy of educational objectives, Krathwohl's affective domain taxonomy, and contemporary theories of value internalization, adapted specifically for the CIPE context. The cognitive stage represents students' acquisition and understanding of ideological concepts and their connection with disciplinary knowledge, essentially the intellectual foundation of ideological education. The affective stage involves the development of emotional resonance, value identification, and attitudinal change ideological content, representing the shift from intellectual recognition to acceptance. The behavioral stage the application of ideological perspectives in practical contexts, decision-making, and professional conduct, representing the observable outcome of successful ideological integration.

Cognitive transfer theory serves as a core theoretical lens for understanding ideological knowledge integrates with disciplinary understanding. According to this perspective, deep CIPE integration occurs when students can transfer ideological concepts and analytical frameworks into their understanding of knowledge, creating interconnected knowledge structures rather than isolated information silos. The TPACK framework further informs our model by emphasizing the intersection of content knowledge, pedagogical knowledge, and, extended for this study, ideological knowledge, suggesting that effective integration requires teachers to navigate the intersections of all three domains. Value internalization models from social psychology provide the theoretical mechanism for understanding how externally presented ideological content becomes meaningful. Following Kelman's classic framework of compliance, identification, and internalization, this study conceptualizes deep integration as moving beyond superficial compliance toward genuine internalization where ideological values become part of the student's personal value system. Educational ecology theory provides the systems-level perspective, conceptualizing CIPE integration as a dynamic system involving multiple interacting elements: teachers, students, content, technology, and institutional context. Integration quality emerges from the interactions of these elements rather than residing in any single element [26].

Four core dimensions of deep integration

Upon the theoretical framework and systematic literature analysis, multi-disciplinary case deconstruction, and expert consultation, this study identifies four core dimensions that constitute deep CIPE integration: political guidance, cultural identity, ethical norms, and practical innovation. These dimensions represent distinct yet interrelated facets of ideological and political elements integrating with disciplinary curricula, collectively providing comprehensive coverage of CIPE objectives while maintaining parsimony for practical assessment applications. Each dimension captures a different aspect of ideological integration, and together they provide a holistic profile of integration quality rather than relying on a single aggregate score. The hierarchical indicator system derived from this multidimensional framework is illustrated in Table 1, which further decomposes each dimension into second-level criteria and corresponding representative indicators, and assigns weights to quantify the relative importance of each dimension.

Table 1. CIPE deep integration indicator system and weights.

Dimension	Weight	Second-level criteria	Representative indicators
Political guidance	0.287	Theoretical awareness Policy cognition Marxist methodology	Correct political orientation Theoretical depth of integration Policy relevance
Cultural identity	0.243	Cultural heritage Cultural confidence Humanistic spirit	Traditional culture integration Cultural self-awareness Humanistic value transmission
Ethical norms	0.265	Professional ethics Social responsibility Moral reasoning	Ethical case integration Social responsibility awareness Moral dilemma analysis

Dimension	Weight	Second-level criteria	Representative indicators
Practical innovation	0.205	Innovative spirit Practical capability Entrepreneurial consciousness	Innovation-oriented activities Real-world problem application Value-driven practice

Political guidance dimension refers to the extent to which courses effectively transmit correct political orientations, theoretical perspectives, and policy awareness within the disciplinary context. This dimension captures how well students develop understanding of socialism with Chinese characteristics, grasp the political implications of their field, and cultivate the ability to analyze professional issues from Marxist perspectives. The cultural identity dimension assesses the integration of Chinese cultural heritage, cultural confidence, and humanistic spirit into professional education, recognizing that ideological education encompasses not only political theory but also cultural transmission and national identity formation. The ethical norms dimension focuses on the integration of professional ethics, social responsibility, and moral reasoning into disciplinary education, addressing how courses prepare students to navigate the ethical dimensions of their profession and understand the social impacts of their work. The practical innovation dimension evaluates how CIPE integration fosters innovative spirit, practical capability, and entrepreneurial consciousness with ideological grounding, reflecting the emphasis on applying ideological perspectives to real-world problem-solving and innovative practice.

Research hypotheses

Based on the theoretical framework and literature review, this study formulates four research hypotheses to guide our empirical investigation. These hypotheses address the validity of our assessment approach, the effectiveness of our support strategies, and the mechanisms through which CIPE integration operates, providing testable propositions that structure our data analysis and interpretation of findings.

Hypothesis 1: The multi-dimensional indicator system incorporating political guidance, cultural identity, ethical norms, and practical innovation demonstrates good construct validity and inter-rater reliability for assessing CIPE deep integration across different disciplines. This hypothesis predicts that our four-dimensional structure will be confirmed through factor analysis and that the indicator system will achieve high consistency when used by different expert evaluators, supporting its use as a standardized assessment framework.

Hypothesis 2: The generative AI-powered multimodal assessment method achieves high agreement with expert human evaluation of CIPE integration quality, outperforming traditional text-based automated assessment approaches. Specifically, this study predicts that the dual-cycle system will produce evaluation scores with intraclass correlation coefficients exceeding 0.80 when compared with expert ratings, and that multimodal data integration will improve assessment accuracy relative to text-only approaches.

Hypothesis 3: The adaptive support strategy framework based on closed-loop optimization significantly improves CIPE integration depth compared with conventional instructional support approaches. This study hypothesizes that courses receiving AI-supported adaptive interventions will show statistically significant improvements in integration scores over time, with effect sizes exceeding those of control groups receiving standard CIPE training.

Hypothesis 4: The effectiveness of CIPE support strategies varies across disciplinary contexts, with humanities and social sciences courses showing greater responsiveness to cognitive-oriented interventions and science, technology, engineering, mathematics (STEM) courses benefiting more from ethics and practical innovation-focused strategies. This hypothesis reflects our theoretical expectation that disciplinary differences moderate intervention effectiveness, supporting the need for adaptive, discipline-specific support approaches.

Methodology

Research design and setting

This study employs a mixed-methods research design combining indicator system construction, technical system development, and quasi-experimental validation. The research proceeds in three sequential phases: First, the construction and validation of the CIPE deep integration evaluation indicator system. Second, the development of the generative AI-powered multimodal assessment system. Third, the implementation and evaluation of support strategies through design-based research and quasi-experimental methods. This multi-phase design allows for systematic development and rigorous validation of the proposed framework, aligning with engineering research paradigms in educational technology that emphasize both theoretical soundness and practical utility. The design enables us to build progressively from foundational indicator development through technical implementation to empirical testing of effectiveness [27].

The research setting is a comprehensive university in eastern China. This study involves six undergraduate courses spanning three disciplinary domains. The six courses contain two humanities courses and two social science courses, namely Chinese Philosophy, Modern Chinese History, Introduction to Law and Introduction to Economics. The remaining two courses belong to STEM, which are Introduction to Artificial Intelligence and Environmental Engineering Principles. These courses were purposefully selected to represent disciplinary variation in how CIPE elements manifest, enabling us to test the disciplinary adaptability of our assessment framework. Each course has an enrollment of 45 to 60 students and is taught by experienced instructors who have received basic CIPE training. Data collection spans one full semester of 16 teaching weeks for each course, encompassing multiple data types: classroom video recordings, teaching text materials, audio transcripts of lectures, student assignment submissions, classroom interaction logs, questionnaire survey data, and expert evaluation scores. Three independent experts in ideological and political education provide baseline and endpoint evaluations for validity comparison. The study received institutional review board approval, and all participating instructors and students provided informed consent prior to data collection.

Indicator system construction

The evaluation indicator system is developed through a multi-stage process combining theoretical analysis, expert consultation, and empirical validation. Following the three-level architecture of objective level, criterion level, and indicator level, this study first derives candidate indicators from the four core dimensions and the cognitive-affective-behavioral three-stage model. This initial pool of 48 candidate indicators is refined through three rounds of Delphi expert consultation involving 15 experts in ideological education, curriculum design, and educational assessment. The Analytical Hierarchy Process is employed to determine indicator weights, with expert pairwise comparison data analyzed to calculate priority weights for each level. Consistency ratios are computed to ensure judgment consistency, with all matrices achieving composite reliability (CR) below 0.10 indicating acceptable consistency. The resulting indicator system comprises 4 first-level dimensions, 12 second-level criteria, and 36 third-level specific indicators, each with operational definitions and scoring anchors. The final weights are: political guidance (0.287), cultural identity (0.243), ethical norms (0.265), and practical innovation (0.205), reflecting the relative importance of each dimension as determined by expert consensus [28].

To address disciplinary differences, this study develops an adaptive quantification algorithm that incorporates discipline characteristic coefficients. These coefficients, derived from disciplinary analysis and expert judgment, adjust indicator weights and scoring benchmarks based on the disciplinary domain. For example, the cultural identity dimension receives a 1.15 weighting coefficient in humanities courses, while the ethical norms dimension receives a 1.20 coefficient in STEM courses, reflecting their relative importance in different disciplinary contexts. This adaptive mechanism ensures that the evaluation system maintains its core structure while providing discipline-appropriate assessment standards, avoiding the one-size-fits-all problem that plagues many existing CIPE evaluation instruments. The algorithm also accounts for course-specific characteristics such as class size, teaching mode, and assessment format, further refining the adaptability of evaluation criteria.

Multimodal data collection and assessment method

A comprehensive multimodal data collection system is deployed to capture CIPE integration across classroom, practical, and online learning scenarios. The system integrates four data categories: textual data (teaching materials, assignment submissions, discussion board posts), audio data (classroom lecture recordings with transcription), video data (classroom recordings for behavioral and emotional analysis), and behavioral log data (learning management system interactions, participation patterns). Data collection is synchronized across modalities to enable temporal alignment for integrated analysis. For multimodal data representation, this study implements a BERT-LSTM hybrid neural network architecture with multi-modal alignment networks. Textual data is processed through a fine-tuned BERT model to generate contextual semantic embeddings. Temporal sequence data is processed through LSTM layers that capture sequential dependencies. The multi-modal alignment network learns cross-modal attention weights, enabling the model to identify correspondences between different data types. The output is a three-dimensional CIPE integration characterization matrix encompassing semantic, temporal, and spatial dimensions [29].

The core of our assessment methodology is a dual-cycle system built upon large language models. A three-layer hierarchical prompt instruction template structure is adopted in this research. The concept mapping layer is used for identifying and mapping implicit ideological elements within disciplinary knowledge, and the situational reasoning layer serves for creating discipline-specific moral dilemma scenarios to evaluate value conflict handling. The value evaluation layer is designed to implement multi-agent, multi-modal value internalization assessment. Low-Rank Adaptation (LoRA) fine-tuning on a domain-specific dataset of CIPE evaluation examples trains the model to align with our indicator system and scoring standards.

A PPO reinforcement learning mechanism constructs a reward model based on multi-dimensional scoring rules, dynamically adjusting knowledge retrieval weights and evaluation strategies based on feedback. The dual-cycle structure operates with an outer cycle refining prompt templates and evaluation criteria based on expert feedback, and an inner cycle optimizing model parameters through reinforcement learning from assessment outcomes. The final assessment output is generated by a Transformer-based multi-task evaluation model that simultaneously scores all 36 indicators while producing explanatory justifications, with comparative judgment matrices validating indicator effectiveness and self-explanatory integration index cloud maps providing visual diagnostic outputs [30-32].

Support strategies and data analysis

The support strategy framework is organized around two primary intervention directions, cognitive reconstruction and behavioral guidance, encompassing three strategy categories: diagnostic prompts, affective motivation, and resource provision. These strategies are designed to be dynamically triggered based on real-time assessment outputs from the evaluation system, creating a closed-loop improvement cycle. Cognitive reconstruction strategies focus on enhancing teachers' understanding and application of CIPE integration principles through discipline-ideology concept network construction tools and metacognitive question generators. Behavioral guidance strategies provide concrete instructional interventions including teaching event intelligence generation, where the system suggests specific learning activities or case studies that would strengthen ideological integration in identified weak areas. The intervention matching algorithm maps assessment results to appropriate strategy types, determining optimal intervention timing, delivery format, and intensity level. The implementation architecture integrates DeepSeek as the primary large language model with application programming interface (API) connections to additional open-source models, providing multi-model redundancy and specialized capabilities for different task types.

Quantitative data analysis employs multiple statistical methods to test the research hypotheses. For indicator system validation, this study conducts confirmatory factor analysis to verify the four-dimensional structure and calculate Cronbach's α coefficients for internal consistency reliability, with inter-rater reliability assessed using intraclass correlation coefficients. For assessment method validation, this study computes Cohen's weighted κ and intraclass correlation coefficients between AI-generated scores and expert ratings, supplemented by Bland-Altman analysis to examine systematic bias. Performance comparisons between multimodal and text-only assessment configurations use paired t-tests on accuracy metrics. For intervention effectiveness evaluation, this study employs mixed-effects Analysis of Variance (ANOVA) models with time as a within-subjects factor and group as a between-subjects factor, testing for interaction effects. Disciplinary moderation effects are examined through three-way interactions including discipline category. Effect sizes are reported as partial eta-squared for ANOVA results and Cohen's d for pairwise comparisons. All statistical analyses are conducted using R with appropriate packages for mixed-effects modeling, with significance thresholds set at $\alpha=0.05$ with Bonferroni correction for multiple comparisons.

Results

Indicator system validation results

The three-round Delphi consultation achieved satisfactory consensus levels among the 15 participating experts. The authority coefficient across all dimensions ranged from 0.78 to 0.86, exceeding the 0.70 threshold for good expert authority. Kendall's coefficient of concordance for the third round reached 0.723 ($p<0.001$), indicating strong agreement among experts regarding indicator importance rankings. The final indicator system retained 36 specific indicators across four first-level dimensions and twelve second-level criteria, with twelve initially proposed indicators removed or merged due to low importance ratings or conceptual redundancy.

Confirmatory factor analysis results supported the four-dimensional structure of the CIPE integration indicator system, with model fit indices all meeting conventional thresholds for acceptable fit: chi-square over df=1.87, comparative fit index (CFI)=0.942 (see Table 2), Tucker-Lewis index (TLI)=0.928, root mean square error of approximation (RMSEA)=0.058, standardized root mean square residual (SRMR)=0.047. Standardized factor loadings for all indicators exceeded 0.62, with most ranging between 0.71 and 0.89, indicating strong indicator-to-construct relationships. Internal consistency reliability was excellent, with Cronbach's α of 0.924 for the overall scale and ranging from 0.847 to 0.893 across the four dimensions. Inter-rater reliability analysis among three independent experts yielded intraclass correlation coefficients of 0.857 for overall integration scores, indicating excellent inter-rater consistency and confirming Hypothesis 1.

Table 2. Summary of research hypothesis testing results.

Hypothesis	Key measure	Result	Conclusion
Hypothesis 1: indicator system validity	CFI model fit; intraclass correlation coefficient (ICC)	CFI=0.942; ICC=0.857	Supported
Hypothesis 2: AI-expert agreement	Cohen's κ; accuracy gain	κ=0.823; +11.4% multimodal	Supported
Hypothesis 3: intervention effectiveness	Time x Group interaction; d	F=32.470, p<0.001; d=1.380	Supported
Hypothesis 4: disciplinary moderation	Three-way interaction	F=3.820, p=0.005; partial support	Partially supported

Assessment model performance

The generative AI assessment model demonstrated strong performance in evaluating CIPE integration quality. Comparison between AI-generated scores and expert ratings across 96 classroom sessions yielded a weighted Cohen's κ of 0.823 for overall integration level classification, indicating substantial to almost perfect agreement. The intraclass correlation coefficient for continuous scores was 0.871, exceeding the 0.80 threshold predicted in Hypothesis 2. Dimension-level agreement varied somewhat: political guidance ($\kappa=0.851$), ethical norms ($\kappa=0.834$), cultural identity ($\kappa=0.798$), and practical innovation ($\kappa=0.776$), with the slightly lower values for practical innovation likely reflecting the greater difficulty of assessing behavioral outcomes from primarily observational data.

The multimodal configuration significantly outperformed text-only assessment, providing empirical support for the value of multi-channel data integration. Paired t-tests revealed that multimodal assessment achieved 11.4% higher accuracy compared with text-only evaluation ($t=4.73$, $df=95$, $p<0.001$), with the most pronounced improvements in dimensions involving affective and behavioral components. Bland-Altman analysis indicated no systematic bias between AI and expert assessments (mean difference =

1.27 on a 100-point scale), with 94.8% of observations falling within the 95% limits of agreement, suggesting acceptable concordance for practical application. The reinforcement learning enhancement contributed measurable improvements in assessment quality, with PPO-based fine-tuning producing a 7.3 percentage point improvement in overall agreement with expert ratings (from 79.8% to 87.1%), along with reductions in evaluation bias and improved interpretability as rated by expert reviewers. These results collectively support Hypothesis 2, confirming that the generative AI-powered multimodal assessment method achieves high agreement with expert evaluation and outperforms traditional text-based approaches.

Intervention effectiveness results

The quasi-experimental evaluation revealed significant improvements in CIPE integration depth for courses receiving the AI-supported adaptive intervention. Mixed-effects ANOVA showed a significant time-by-group interaction effect for overall integration scores $F(2, 188) = 32.470$, $p<0.001$, partial eta-squared = 0.257, indicating differential improvement trajectories between intervention and control groups (see Table 3). Post-hoc analyses revealed that the intervention group improved by 23.7% from pre-test to post-test (M_{pre} = 58.3, SD=8.7; M_{post} = 72.1, SD=7.9), compared with a 6.2% improvement in the control group (M_{pre} = 57.9, SD=9.1; M_{post} = 61.5, SD=8.4). The between-group effect size at post-test was large (Cohen's d = 1.38), providing strong support for Hypothesis 3.

Dimension-level analyses revealed differential improvement patterns across the four CIPE dimensions. The intervention group showed the largest gains in ethical norms (28.3% improvement) and practical innovation (26.1% improvement), followed by political guidance (22.4%) and cultural identity (19.8%). Control group improvements were modest across all dimensions, ranging from 4.1% to 7.9%. Analysis of strategy usage patterns indicated that teachers most frequently utilized the diagnostic prompt tools (68.0% of system interactions) and concept network visualizations (57.0%), while affective motivation strategies had the highest per-use impact on subsequent integration scores despite being less frequently accessed. The disciplinary moderation analysis partially supported Hypothesis 4, revealing a significant three-way interaction of time, group, and discipline $F(4, 188) = 3.820$, $p=0.005$, partial eta-squared = 0.075. Humanities courses showed the greatest responsiveness to the intervention overall (31.2% improvement), followed by social sciences (22.8%) and STEM courses (17.1%). Contrary to our prediction, STEM courses showed strong improvements in the ethical norms dimension (32.7%) but benefited substantially from cognitive-oriented interventions that helped identify ideological connections within technical content.

Table 3. Intervention effects on CIPE integration by dimension and discipline.

Category	Intervention group improvement (%)	Control group improvement (%)	Effect size (Cohen's d)
Overall	23.7	6.2	1.38
Political guidance	22.4	5.8	1.25
Cultural identity	19.8	4.1	1.07
Ethical norms	28.3	7.9	1.52
Practical innovation	26.1	6.5	1.41

Integration index cloud map results

The CIPE integration index cloud maps generated by the system provided rich diagnostic information that was well-received by participating teachers. The visualization successfully captured dynamic patterns of integration across different lesson phases, topic areas, and activity types. Temporal analysis revealed characteristic patterns such as higher integration density during introductory and concluding lesson segments, with lower integration during middle sections focused on technical content delivery. These patterns were consistent across disciplines but varied in magnitude, with humanities courses showing more uniformly distributed integration and STEM courses showing more concentrated bursts of ideological content.

Cross-dimensional correlation patterns visible in the cloud maps revealed interesting insights about integration dynamics. Political guidance and cultural identity dimensions showed the strongest co-occurrence (r=0.74), suggesting that these dimensions are often addressed together through similar instructional approaches. Ethical norms and practical innovation showed moderate co-occurrence (r=0.58), appearing more frequently in applied and case-based learning activities. The spatial distribution visualization successfully identified cold spots, specific course topics or activity types where integration was consistently weak, which proved valuable for targeting interventions. Teacher survey responses indicated that 83.0% of participants found the cloud map visualizations helpful for understanding their CIPE integration patterns and identifying improvement areas.

Discussion and implication

Key findings discussion

The results of this study provide robust empirical support for the proposed generative AI-supported CIPE assessment and intervention framework. The indicator system validation results confirm that CIPE deep integration can be meaningfully operationalized along four core dimensions: political guidance, cultural identity, ethical norms, and practical innovation, with good psychometric properties. This finding advances the field by moving beyond broad, undifferentiated conceptions of CIPE toward a structured, multi-faceted understanding that captures both the content domains and depth of integration. The strong inter-rater reliability suggests that with appropriate training and clear operational definitions, human evaluators can achieve consistent assessments using this framework, providing a solid foundation for both manual and AI-based evaluation approaches.

The assessment model performance results are particularly noteworthy, demonstrating that generative AI with appropriate prompt engineering and multimodal data integration can approach human expert levels of CIPE evaluation accuracy. The κ value of 0.823 indicates substantial agreement that meets or exceeds typical inter-rater agreement among human evaluators in educational assessment contexts. The superiority of multimodal over text-only assessment underscores the importance of capturing the richness of classroom interactions, including paralinguistic cues, behavioral patterns, and emotional dynamics, that text-based approaches miss, aligning with broader trends in learning analytics research. The reinforcement learning enhancement's contribution to model improvement highlights the value of iterative, feedback-driven approaches to AI assessment in educational contexts. The 7.3 percentage point improvement from PPO fine-tuning demonstrates that initial prompt engineering alone, while effective, can be substantially enhanced through mechanisms that incorporate expert feedback into the model's evaluation strategies. This dual-cycle approach offers a generalizable methodology for adapting LLMs to specialized educational assessment tasks where labeled training data may be limited but expert feedback is available.

The intervention effectiveness results provide strong evidence that AI-supported adaptive strategies can meaningfully improve CIPE integration quality. The 23.7% improvement in the intervention group, compared with 6.2% in controls, represents a substantial effect that has practical significance for educational practice. Importantly, the improvement was observed across all disciplinary domains, though with varying magnitudes, suggesting that the approach has broad applicability. The finding that humanities courses showed the greatest improvement likely reflects the more natural alignment between humanistic content and ideological education, making integration opportunities easier to identify and implement with appropriate scaffolding. STEM courses, while showing smaller relative gains, still achieved meaningful improvements, demonstrating that the adaptive approach can effectively support even disciplines where CIPE integration is traditionally more challenging. The disciplinary moderation findings reinforce that context adaptation matters in educational interventions, supporting ecological perspectives that emphasize fit between interventions and their implementation contexts.

Theoretical and practical implications

This study makes several theoretical contributions to the fields of curriculum ideological education and educational technology.

First, it elaborates a comprehensive theoretical framework for CIPE deep integration that synthesizes cognitive transfer theory, value internalization models, and educational ecology perspectives. The four-dimensional structure provides a theoretically grounded taxonomy for understanding how ideological elements manifest across different aspects of disciplinary education, moving beyond vague conceptualizations toward a structured framework that can guide both research and practice.

Second, the cognitive-affective-behavioral three-stage model adapted for CIPE context contributes to understanding the developmental progression of ideological integration. The assessment framework's ability to distinguish between different stages of value internalization addresses a critical gap in current CIPE research, which often treats integration as a binary quality rather than a developmental continuum.

Third, the study advances theoretical understanding of how generative AI can support assessment in values-laden educational domains. The dual-cycle model provides a theoretical mechanism for how AI systems can learn to perform complex evaluative judgments in domains where objective scoring criteria are elusive and expert judgment is required.

Fourth, the adaptive intervention framework contributes to theories of instructional support for curriculum innovation. It operates design-based research principles within an AI-enhanced system and demonstrates how continuous assessment data can drive personalized, context-appropriate instructional improvements.

For educational practitioners, this study provides several actionable implications. The validated indicator system offers a structured framework that teachers and instructional designers can use for self-assessment and curriculum planning. The discipline-adaptive coefficients help contextualize these standards for different fields, acknowledging that what constitutes good integration varies across disciplines. The AI assessment system provides a practical tool for ongoing monitoring and diagnosis of CIPE integration quality, offering continuous feedback with minimal overhead. The integration index cloud maps offer intuitive visualizations that make complex assessment results accessible and actionable. The adaptive support strategy framework provides concrete, evidence-based intervention approaches dynamically matched to identified needs, increasing the efficiency and effectiveness of support efforts. For educational administrators and policymakers, the substantial effect sizes demonstrate that investing in intelligent assessment and support systems can yield meaningful improvements in CIPE implementation quality.

Conclusion

This study sets out to develop and validate a generative AI-supported framework for assessing and enhancing the deep integration of curriculum ideological and political education. Through systematic indicator construction, technical system development, and quasi-experimental evaluation, this study has achieved the three research objectives established at the outset. First, this study constructs and validates a fine-grained, discipline-adaptive evaluation indicator system encompassing four core dimensions: political guidance, cultural identity, ethical norms, and practical innovation, demonstrating strong construct validity and inter-rater reliability across disciplinary contexts. Second, this study develops a generative AI-powered multimodal assessment methodology that achieves high agreement with expert human evaluation (κ=0.823), significantly outperforming text-only approaches through the integration of multi-channel classroom data. Third, this study designs and validates an adaptive support strategy framework that produces substantial improvements in CIPE integration depth (23.7% improvement vs. 6.2% in controls) through intelligent diagnosis, targeted intervention, and evidence-based optimization cycles.

The findings collectively demonstrate that generative AI technology, when combined with sound educational theory and carefully designed assessment frameworks, can effectively support the evaluation and enhancement of curriculum ideological and political education. The technical pathway of multimodal perception, generative reasoning, and dynamic optimization provides a viable approach for addressing the assessment challenges inherent in values-laden educational domains, where nuanced judgment and contextual understanding are required. The closed-loop support strategy model offers a practical mechanism for translating assessment insights into instructional improvement, creating closed-loop quality enhancement systems. This research contributes to the broader endeavor of educational digitalization by demonstrating how advanced AI technologies can be responsibly and effectively applied to ideological and political education, a domain where technological interventions must be carefully calibrated to educational objectives and ethical considerations.

The methodological approaches developed here include the prompt engineering dual-cycle assessment system and the adaptive intervention framework. These approaches may have applicability to other educational domains involving complex, value-laden assessment challenges, such as ethics education, character development, and professional competency evaluation.

Limitations and prospects

Several limitations of this study should be acknowledged when interpreting the findings. First, the sample is limited to six courses at a single university, which may restrict the generalizability of the results. The participating institution is a comprehensive university in eastern China with relatively strong CIPE implementation foundations, and the findings may not fully generalize to institutions with different characteristics, resource levels, or regional contexts. Future research with larger and more diverse samples across multiple institutions would strengthen confidence in the framework's general applicability.

Second, the study duration of one semester provides only medium-term evidence of intervention effectiveness. The long-term sustainability of improvements and their impact on student outcomes, such as value internalization, behavioral change, and long-term development, were not directly assessed. CIPE aims for deep, enduring value formation that may require extended periods to manifest, and shorter-term improvements in instructional integration quality may not fully translate to long-term student development outcomes.

Third, while the AI assessment system achieves strong agreement with human experts, it is not immune to potential biases inherent in large language models, including those related to training data distributions and value alignment. The reinforcement learning mechanism mitigates but does not eliminate these concerns. Ongoing monitoring, human oversight, and periodic recalibration with expert input remain essential for responsible deployment.

Fourth, the study's focus on classroom-level assessment and intervention does not address institutional-level factors that influence CIPE implementation quality, such as policy frameworks, incentive structures, teacher training programs, and organizational culture.

Several promising directions for future research emerge from this work. First, expanding the assessment framework to incorporate student outcome measures would create a more comprehensive evaluation system. This system connects instructional integration quality with student learning and development outcomes, including measures of value internalization, moral reasoning development, and ideological identity formation.

Second, advancing the technical capabilities of the assessment system through improved multimodal fusion algorithms, enhanced emotion recognition capabilities, and more sophisticated natural language understanding of ideological content would further improve assessment accuracy and granularity.

Third, developing more sophisticated adaptive intervention algorithms that can learn from intervention effectiveness data to personalize recommendations at the individual teacher level would enhance the support system's precision, moving from adaptive to truly personalized support.

Fourth, investigating the scalability and sustainability of AI-supported CIPE systems in real-world institutional deployment contexts would address important practical questions about implementation barriers, teacher acceptance, integration with existing technology ecosystems, and cost-effectiveness.

Finally, exploring the ethical, social, and policy implications of AI-enhanced ideological and political education is essential as these technologies become more prevalent. Important questions about data privacy, algorithmic fairness, teacher autonomy, and the appropriate role of AI in values education require careful interdisciplinary examination. Ethical guidelines and governance frameworks are crucial for ensuring responsible deployment that serves educational objectives rather than unintended consequences.

Funding

This work was supported by the Provincial Undergraduate Training Program on Innovation and Entrepreneurship (Grant No. S202510345090).

Acknowledgements

The authors would like to show sincere thanks to those technicians who have contributed to this research.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] Wang, J. (2021) Analysis of challenges and countermeasures of ideological and political education in colleges and universities in the new era. *Journal of Higher Education Research*, 2(6), 35.

[2] Cheng, P. (2022) Evaluation method of ideological and political classroom teaching quality based on analytic hierarchy process. *Scientific Programming*, 2022(1), 6554084.

[3] Yan, L., Greiff, S., Teuber, Z., Gašević, D. (2024) Promises and challenges of generative artificial intelligence for human learning. *Nature Human Behaviour*, 8(10), 1839-1850.

[4] Miseliunaite, B., Kliziene, I., Cibulskas, G. (2022) Can holistic education solve the world's problems: a systematic literature review. *Sustainability*, 14(15), 9737.

[5] Du, J., Chen, X., Dou, Y. (2024) Evaluating the practical effectiveness of college counselors' ideological and political education using big data video streaming. *Wireless Networks*, 30(1), 1-15.

[6] Yang, Y., Long, Y., Sun, D., Van Aalst, J., Cheng, S. (2020) Fostering students' creativity via educational robotics: an investigation of teachers' pedagogical practices based on teacher interviews. *British Journal of Educational Technology*, 51(5), 1826-1842.

[7] Liu, W., He, C. (2022) Curriculum-based ideological and political education: research focuses and evolution. *International Education Studies*, 15(5), 28-35.

[8] Ruan, X., Yang, X. (2024) Integrating ideological and political education into curriculum development in the new era: perspectives from English as a foreign language teaching in universities of China. *Open Journal of Modern Linguistics*, 14(03), 344-352.

[9] Wu, J., Wang, X. (2022) Research on the mode of specialized and integrated education in colleges and universities based on the perspective of curriculum ideology and politics. *International Journal of Education and Humanities*, 3(1), 66-69.

[10] Chiva-Bartoll, O., Fernández-Rio, J. (2022) Advocating for service-learning as a pedagogical model in physical education: towards an activist and transformative approach. *Physical Education and Sport Pedagogy*, 27(5), 545-558.

[11] Leary, M. P., Sherlock, L. A. (2020) Service-learning or internship: a mixed-methods evaluation of experiential learning pedagogies. *Education Research International*, 2020(1), 1683270.

[12] Ren, W., Wang, R., Mohamad, S. N. A., Xie, Y., Chen, L., Ning, H., Shi, J. (2024) A quantitative analysis of the influence of ideological and political education on students' learning satisfaction. *Journal of Infrastructure, Policy and Development*, 8(1), 1-14.

[13] Zhang, X., Zhang, P., Shen, Y., Liu, M., Wang, Q., Gašević, D., Fan, Y. (2024) A systematic literature review of empirical research on applying generative artificial intelligence in education. *Frontiers of Digital Education*, 1(3), 223-245.

[14] AlAli, R., Wardat, Y. (2024) Opportunities and challenges of integrating generative artificial intelligence in education. *International Journal of Religion*, 5(7), 784-793.

[15] Alasadi, E. A., Baiz, C. R. (2023) Generative AI in education and research: opportunities, concerns, and solutions. *Journal of Chemical Education*, 100(8), 2965-2971.

[16] Ren, M., Chen, N., Qiu, H. (2023) Human-machine collaborative decision-making: an evolutionary roadmap based on cognitive intelligence. *International Journal of Social Robotics*, 15(7), 1101-1114.

[17] Batista, J., Mesquita, A., Carnaz, G. (2024) Generative AI and higher education: trends, challenges, and future directions from a systematic literature review. *Information*, 15(11), 676.

[18] Ng, D. T. K., Chan, E. K. C., Lo, C. K. (2025) Opportunities, challenges and school strategies for integrating generative AI in education. *Computers and Education: Artificial Intelligence*, 8, 100373.

[19] Kurtz, G., Amzalag, M., Shaked, N., Zaguri, Y., Kohen-Vacs, D., Gal, E., Zailer, G., Barak-Medina, E. (2024) Strategies for integrating generative AI into higher education: Navigating challenges and leveraging opportunities. *Education Sciences*, 14(5), 503.

[20] Vo, N. N., Vu, Q. T., Vu, N. H., Vu, T. A., Mach, B. D., Xu, G. (2022) Domain-specific NLP system to support learning path and curriculum design at tech universities. *Computers and Education: Artificial Intelligence*, 3, 100042.

[21] Luo, Y., Yang, Y. (2024) Large language model and domain-specific model collaboration for smart education. *Frontiers of Information Technology & Electronic Engineering*, 25(3), 333-341.

[22] Mohammadi, M., Tajik, E., Martinez-Maldonado, R., Sadiq, S., Tomaszewski, W., Khosravi, H. (2025) Artificial intelligence in multimodal learning analytics: a systematic. *Computers and Education*, 8, 100426.

[23] Chango, W., Lara, J. A., Cerezo, R., Romero, C. (2022) A review on data fusion in multimodal learning analytics and educational data mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 12(4), e1458.

[24] Guettala, M., Bourekkache, S., Kazar, O., Harous, S. (2024) Generative artificial intelligence in education: Advancing adaptive and personalized learning. *Acta Informatica Pragensia*, 13(3), 460-489.

[25] Mohamed, F. N., Azhar, J., Yasmeen, S., Hussain, I., Khawar, M. (2025) Generative artificial intelligence and personalized learning environment: challenges and opportunities. *Southern Journal of Computer Science*, 1(01), 1-36.

[26] Yu, J., Song, J., Lu, Y. (2025) Harnessing generative artificial intelligence to construct multimodal resources for Chinese character learning. *Systems*, 13(8), 692.

[27] Lu, X. (2024) The internal mechanisms and practical paths for leveraging generative artificial intelligence to empower high-quality development in university ideological and political education. *Adult and Higher Education*, 6(8), 80-86.

[28] Hoadley, C., Campos, F. C. (2022) Design-based research: What it is and why it matters to studying online learning. *Educational Psychologist*, 57(3), 207-220.

[29] Li, Y., Cheng, H., Qin, Q. (2025) Evaluations and improvement methods of deep learning ability in blended learning. *International Journal of e-Collaboration (IJeC)*, 21(1), 1-17.

[30] Zhang, B., Velmayil, V., Sivakumar, V. (2023) A deep learning model for innovative evaluation of ideological and political learning. *Progress in Artificial Intelligence*, 12(2), 119-131.

[31] Kooli, C., Yusuf, N. (2025) Transforming educational assessment: insights into the use of ChatGPT and large language models in grading. *International Journal of Human-Computer Interaction*, 41(5), 3388-3399.

[32] Fagbohun, O., Iduwe, N. P., Abdullahi, M., Ifaturoti, A., Nwanna, O. M. (2024) Beyond traditional assessment: Exploring the impact of large language models on grading practices. *Journal of Artificial Intelligence, Machine Learning and Data Science*, 2(1), 1-8.