Exploration on Teaching Reform for the" Database Principles and Applications " Course Based on CDIO-OBE Concept

Qing Qu [1, a*], Yuxin Li [2, b]

[1,2] Xi'an Technology and Business College, Xi'an, shaanxi, China

[a] 1451129680@qq.com, [b] 2858436721@qq.com

Abstract. Aiming at the problems existing in the traditional teaching of "Principles and Applications of Database Systems" for software engineering majors in private application-oriented undergraduate universities—such as overemphasis on theory while neglecting practice, significant differences in students' foundational knowledge, and a single evaluation method—this study integrated the CDIO (Conceive-Design-Implement-Operate) engineering education model with the OBE (Outcome-Based Education) concept, and implemented a systematic teaching reform using the "Student Performance Management System" as the core teaching case. Following the CDIO process of "Conceive-Design-Implement-Operate" and relying on SQL Server as the database tool, student-centered project-based teaching was carried out. Adhering to the logic of "reverse design and forward implementation", a three-dimensional objective system covering knowledge, competence, and quality was reconstructed. Progressive project-based teaching was adopted, along with a differentiated strategy of "stratified grouping + flexible tasks", and a multi-dimensional assessment system consisting of "process evaluation (60%) + summative evaluation (40%)" was established. Teaching practice shows that this model effectively stimulates students' learning initiative, and significantly enhances their engineering practice and teamwork abilities. The CDIO-OBE integrated model can effectively improve the teaching quality of application-oriented courses in private universities, providing a referable path for the reform of similar courses.

Keywords: CDIO； OBE; Database Principles; Applications； Teaching Reform； Diversified Assessment.

1. Introduction Outcome-Based Education (OBE) is an educational model designed to cultivate the knowledge, skills, and qualities students need to achieve success. CDIO (Conceive-Design-Implement-Operate) represents a new engineering education philosophy and implementation system focused on student engineering practice, serving as a highly effective engineering education model for enhancing engineers' capabilities [1].

With the rapid advancement of information technology, enterprises' demand for the capabilities of fresh graduates in software engineering has shifted from single-skill proficiency to comprehensive competence. Private universities serve as vital training grounds for applied talent. Reforming curricula to enhance graduates' job readiness is crucial, enabling them to handle database development and maintenance tasks for small and medium-sized enterprises. However, traditional teaching models face certain challenges in private university settings. Disconnect between theoretical instruction and practical application: Classroom focus often centers on abstract concepts like relational algebra and normalization theory, while lab sessions typically involve repetitive, verification-based SQL operations. The absence of comprehensive, project-based learning throughout the curriculum hinders students' ability to translate theoretical knowledge into practical engineering skills. Traditional teaching methods often rely heavily on instructor-led lectures with limited interaction, resulting in low student engagement and diminished interest—particularly in the more theoretical and potentially tedious aspects of database theory. This approach can foster a sense of intimidation among students. Furthermore, conventional assessment methods predominantly emphasize knowledge recall, which may not align with the course's learning objectives, assessment priorities, and engineering characteristics. Neglecting students' varying foundational skills: Students at private universities possess uneven computer literacy, making it difficult for standardized teaching pacing and content to accommodate learners at different proficiency levels. Relying on a single evaluation method: Over dependence on final written exams overlooks the assessment of teamwork, innovative thinking, and engineering practice skills during the learning process [2-3].

To address these issues, scholars worldwide are actively exploring the application of CDIO and OBE principles in engineering education. Multiple institutions have achieved a shift from “teacher-centered” to “student-centered” education through OBE-driven curriculum restructuring [4-6]. However, current research predominantly focuses on applying single methodologies, with limited systematic exploration of the deep integration of CDIO and OBE. Particularly in database principles and applications courses at private universities, tiered instructional design addressing student skill disparities and project-based assessment systems remain underdeveloped.

Based on the CDIO-OBE integration concept, the innovations in the Database Principles and Applications course are as follows:

(1) Tiered Instructional Design: To address variations in students' foundational abilities, a tiered in-class assessment mechanism (covering basic and advanced levels) has been designed to facilitate skill enhancement.

(2) Project-Based Practice Integration: Using a “Student Grade Management System” as the core case study, knowledge points such as E-R diagram design and normalization are integrated into the four CDIO phases, enhancing knowledge coherence and practical applicability.

(3) Diversified Evaluation System: A quantitative framework combining “Formative Assessment (60%) + Summative Assessment (40%)” is established to comprehensively evaluate student performance across all project stages.

2. Current Status and Analysis of the “Database Principles and Applications” Course

2.1 Course Positioning and Characteristics. “Database Principles and Applications” serves as a core course in software engineering programs at private institutions, occupying a pivotal bridging role. It not only connects foundational courses like “Programming Fundamentals” with subsequent courses such as “Web Development” and “Big Data Applications,” but also functions as a key vehicle for cultivating students' abilities in data modeling, SQL programming, and system optimization [7].The core characteristics of this course are reflected in: Theoretical Level: Relational algebra and normalization theory form the mathematical foundation of database design, requiring students to possess strong logical abstraction abilities. Practical Level: Students must master tools like SQL Server to complete end-to-end training—from requirements analysis and E-R modeling to SQL implementation and system optimization—cultivating an engineering mindset [8]. However, this dual emphasis on theory and practice also increases teaching complexity: abstract concepts can intimidate students, while practical sessions lacking systematic project guidance risk becoming fragmented operational drills.

2.2 Analysis of the Teaching Audience. For the 2023 cohort of the Software Engineering program at this institution, comprising 4 classes totaling 139 students. Along with comparable domestic universities, revealed that students at private institutions typically exhibit “strong practical interest but weak theoretical foundations.” Their strengths include high enthusiasm for hands-on activities and the ability to rapidly acquire foundational skills through case-based learning. Weaknesses include poor mathematical foundations, with approximately 40% of students struggling in areas like function dependency derivation and relational algebra operations; limited teamwork experience; and frequent imbalances in task distribution during project practice. These characteristics necessitate a teaching model that integrates differentiated instruction with systematic development of team collaboration skills.

2.3 Necessity of Reform. To address these issues, the integrated concept of “Outcomes-Based Education (OBE) + Capability Development through Industry-Oriented Education (CDIO)” is introduced. Its necessity manifests in three key aspects:(1) Aligning with Industry's Full-Process Competency Demands: Enterprises require database professionals to possess end-to-end capabilities spanning “requirements analysis-design-implementation-operation and maintenance.” Traditional teaching emphasizes the “implementation” phase, whereas CDIO-OBE employs real projects to integrate all four stages, enabling students to experience the complete workflow. (2) Addressing

Varied Student Foundations: Students at private institutions exhibit significant disparities in foundational knowledge. Tiered task design enables participation for learners at all levels, achieving "teaching tailored to individual aptitude." Simultaneously, during group collaboration, instructors can provide personalized guidance based on each student's foundation and abilities, helping them successfully complete tasks while enhancing their sense of learning achievement and confidence. (3) Implementing holistic education goals for new engineering disciplines: Centered on students, CDIO-OBE emphasizes balanced development of competencies and professional ethics. Through group collaboration and project presentations, it cultivates not only specialized knowledge and skills but also enhances communication abilities and professional conduct. These elements align with the new engineering discipline's trinity of goals: value shaping, competency development, and knowledge transmission.

3 Design of Database Course Teaching Reform Based on CDIO-OBE

3.1 Integration of CDIO-OBE Principles and Overall Framework. Although CDIO and OBE principles differ in emphasis, they exhibit strong complementary integration, forming an organic whole that jointly advances the in-depth development of educational reform. The OBE philosophy provides CDIO with a clear goal orientation. During CDIO implementation, OBE helps educators define the ultimate competency objectives students should achieve—that is, "what competencies to cultivate." By thoroughly analyzing industry demands and occupational standards, specific competencies students must possess are identified, giving CDIO projects a clear direction and purpose. CDIO, in turn, provides a practical implementation pathway for OBE. Through the full project lifecycle of "Conceive-Design-Implement-Operate," CDIO translates the competency goals from the OBE philosophy into concrete teaching practices. At each project stage, students progressively master and enhance required competencies through hands-on practice and teamwork. This facilitates the transition from theoretical knowledge to practical skills while simultaneously developing their engineering practice and innovation capabilities. Such experiences validate the attainment of competency goals established under the OBE framework [9-10].

The integrated teaching reform model combining Outcome-Based Education (OBE) and Conceptualization-Design-Implementation-Operation (CDIO) principles is illustrated in Figure 1.

3.2Backward Design: Three-Dimensional Instructional Objectives Based on OBE. Based on enterprise research, the three-dimensional course objectives are designed through backward design:

Knowledge Objectives:

(1) Master fundamental database concepts and principles.

(2) Proficiently operate SQL Server tools and T-SQL programming. Independently complete database creation, table design, data queries, and related operations.

(3) Master the complete database design process from requirements analysis and E-R diagram design to physical implementation.

Competency Objectives: (1) Foundational Competencies: Independently write SQL statements (DQL queries, DML operations, DDL definitions). Achieve ≥90% pass rate on foundational SQL assessments. (2) Advanced Competencies: Master SQL writing techniques. Design E-R diagrams and perform normalization optimization based on system requirements. Translate real-world entities and relationships into conceptual database models. (3) Comprehensive Competency: Collaborate in a team to complete the database design and implementation for a "Student Grade Management System."

Attitude Objectives: (1) Demonstrate teamwork: Learn to collaborate with others in group projects, respect peers' opinions and suggestions, and jointly accomplish project tasks to enhance teamwork skills. (2) Cultivate professional ethics: Adhere to coding standards, maintain complete documentation, develop rigorous engineering thinking, approach problems from an engineering perspective, and prioritize system design, implementation, and maintenance. (3) Cultivating a habit of self-directed learning in new technologies, learning to summarize lessons, continuously improving design and implementation solutions, and enhancing professional competence and overall quality.The interrelationship between the key components is illustrated in Figure 1.

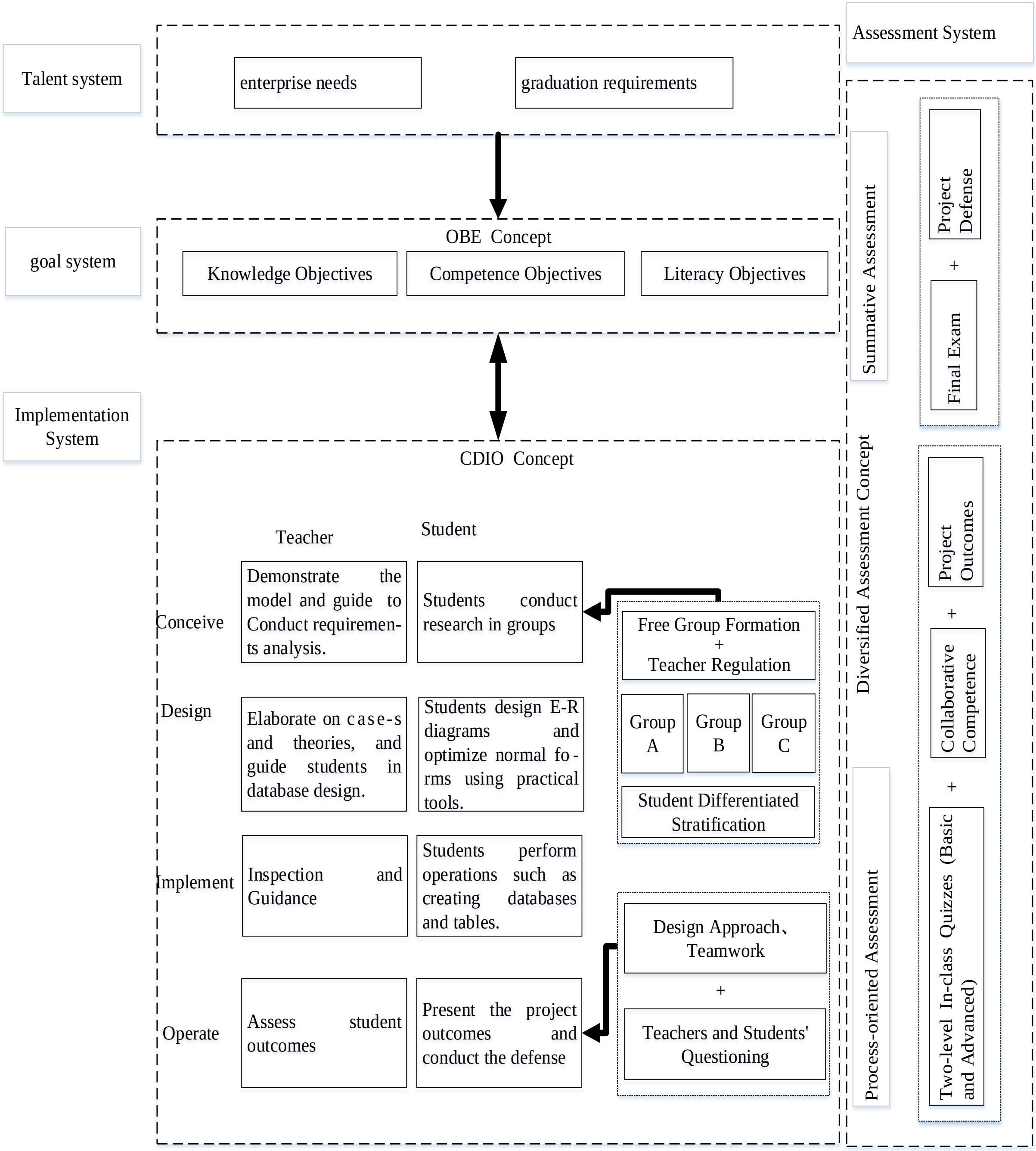

Figure 1: CDIO-OBE Integrated Teaching Reform Framework Diagram

3.3 CDIO-based Progressive Reconstruction of Teaching Content. Centered on the “Student Grade Management System” project, teaching content is decomposed according to the four CDIO stages into progressive tiers: “Foundation-Intermediate-Comprehensive-Advanced.” This approach accommodates varying student foundations while fully leveraging SQL Server tools to achieve a “theory-practice” closed loop. Tool application details for each stage are as follows:

3.4 Implementation of Differentiated Teaching and Process - oriented Assessment Strategy.

(1) Design of Student Stratification and Flexible Tasks. Based on students' prior coursework performance and varying foundational knowledge, a “stratified grouping + flexible tasks” strategy is adopted. Students are divided into three tiers for differentiated instruction:The specific strategies for curriculum design and learner grouping are detailed as illustrated in Table 1 (CDIO-based Progressive Teaching Content Design) and Table 2 (Student Stratification Based on the Implementation of Differentiated Instruction).

Table 1. CDIO-based Progressive Teaching Content Design

Teaching Stage	CDIO Phase	Core Applications of SQL Server Tools	Students' Main Operational Tasks
Basic Stage	Conceive	Create databases via SSMS graphical interface and T-SQL statements; Define table field types using Table Designer; Add primary keys and check constraints with Constraint Manager.	Independently complete database creation and table structure design, and submit screenshots; Write CREATE TABLE statements with T-SQL (compare differences between graphical and code operations); Insert test data.
Advanced Stage	Design	Draw system E-R diagrams; Configure foreign key relationships in Table Designer; Create "score statistics views" via View Designer; Perform multi-table join queries.	Independently draw E-R diagrams and correct foreign key association errors; Write and query views using T-SQL; Complete normalization determination to avoid data redundancy.
Comprehensive Stage	Implement	Develop core system functions; Write triggers to implement business logic.	Develop core functions; Write triggers and test trigger functions.
Enhancement Stage	Operate	Perform user permission management; Prepare for project defense.	Project defense: demonstrate system functions, explain problem-solving processes, and propose improvement directions.

Table 2: Student Stratification Based on the Implementation of Differentiated Instruction

Stratification	Differences in Core Tasks
Group A (Weak Foundation)	Focus on consolidating basic skills. Tasks: Complete table creation and single-table queries with reference to templates; imitate and write multi-table join queries.
Group B (Medium Foundation)	Focus on skill application and transfer. Tasks: Independently complete table creation and multi-table queries; draw preliminary E-R diagrams; independently develop simple stored procedures.
Group C (Strong Foundation)	Focus on system design and optimization. Tasks: Independently complete database design, including E-R diagram optimization and normalization determination; develop complex triggers.

(2) Design and Implementation Effect of In-class Quizzes.

To assess learning outcomes in real time, two in-class quizzes closely integrated with projects are designed:

Quiz 1 (Basic SQL) Core Topics: Create a “Student Table” and insert data; Query “Names and Student IDs of Male Students in Class of 2024”; Update “Zhang San's Grade to 2023”; Delete “Course with ID C001.”
Quiz 2 (Advanced SQL): Core topics: Multi-table join query for “student names, course names, and scores”; subquery to “identify student IDs and scores exceeding the average”; create a view for “Grade 2024 student performance statistics (including average scores)”.
(3) Group Collaboration and Role Rotation.
Employ a “free grouping + teacher fine-tuning” approach to ensure every student experiences the full CDIO process and develops comprehensive skills.
Self-Formation Phase: Groups of 3-4 students, ensuring each includes A, B, and C-level students to achieve “stronger students guiding weaker ones.”
Faculty Adjustment Phase: Groups are restructured to ultimately meet the composition of “1 Group C student + 1-2 Group B students + 1 Group A student.”
Role Rotation Mechanism: Clearly defined roles are rotated to ensure each student participates in the entire process, covering the full cycle of “Requirement Analysis - Design - Development - Testing.”
3.5 Establishing a Diversified Assessment System. Moving beyond the “single final exam” model, we implement a “formative assessment (60%) + summative assessment (40%)” system that balances “knowledge acquisition, skill development, and character cultivation”:

Table 3: Diversified Assessment System

Evaluation Dimension	Evaluation Content and Method	Evaluation Subject	Weight
Process-oriented	Two in-class quizzes; phased project outcomes (requirements documents, E-R diagrams, SQL code); group collaboration performance (contribution degree, communication ability).	Teachers, group leaders	60%
Summative	Final written exam (focusing on comprehensive application); final project defense (system functions, design approach).	Teachers	40%

4 Teaching Reform Practice: CDIO-OBE-Based Instructional Implementation Process

4.1 CDIO Four-Phase Implementation Driven by Projects. Using the “Student Grade Management System” as the vehicle, fully implement the CDIO engineering education model.
(1) Conception Phase (Conceive): Demand-Driven, Goal Clarification.
Task Objective: At the course outset, instructors demonstrate the “Student Grade Management System” prototype through hands-on operation, showcasing system functions such as student information entry, query, and modification; grade input, statistics, and analysis; and report generation. This provides students with an intuitive understanding of the system, its practical application scenarios, and value, stimulating learning interest and motivation. Instructors guide students to analyze the system's functional requirements, prompting them to consider implementation methods and data support for each module. This introduces the course content and lays the foundation for subsequent learning activities.
Implementation Steps: Instructors provide a “Requirement Research Template.” Students conduct interviews in groups, engaging with faculty and peers to understand functional requirements and expectations for the “Student Grade Management System.” Through group discussions, they consolidate these requirements into a “Requirement Specification Document.” During drafting, students must clearly define the system's functional and performance requirements, providing explicit guidance for subsequent design and implementation phases.
(2) Design Phase (Design): Entity-Relationship Diagrams and Normal Form Optimization, Aligned with Job Design Competencies.

Task Objective: Transform requirements into conceptual and logical models, identify and optimize normal forms to ensure data integrity.

Implementation Steps: Teams use tools like Visio to draw E-R diagrams, clarifying relationships between entities; instructors use specific cases to help students master methods for determining 1NF-3NF. After explaining normalization theory, instructors guide students through normalization analysis. Through group discussions, students complete normalized relational schema designs and submit a Database Design Report.。

(3) Implementation Phase（Implement）: Layered Practice to Strengthen Skills.

Task Objective: Implement database designs in SQL Server, develop core functionalities, and complete testing.

Implementation Steps: Basic operations: Write DDL statements to create databases and tables, set primary keys, foreign keys, and constraints; Advanced operations: Write views, stored procedures, and triggers; Functional testing: Develop test cases to validate functionality and data integrity.

Instructors provide on-site guidance and deliver focused lectures on common challenges like "errors in multi-table join queries."。

(4)Operation Phase（Operate）: Present Project Outcomes.

Task Objectives: Optimize database performance, showcase project deliverables, summarize lessons learned, and conduct a final presentation.

Implementation Steps: Comprehensively test the functionality of the "Student Grade Management System." Develop test cases to verify whether the system correctly handles concurrent operations, ensuring data consistency and integrity.

Each group presents for 15 minutes. During the defense, emphasize explaining design concepts—such as E-R diagram design, normalization principles, trigger application, and team collaboration processes—while answering questions from instructors and peers. Defense scores are incorporated into the final evaluation.

4.2 Dynamic Feedback Mechanism. At the conclusion of each phase, instructors publish a Project Progress Report providing detailed analysis of each group's performance in areas such as E-R diagram design and SQL efficiency. The report highlights issues in E-R diagram design—including entity relationship errors and attribute redundancy—and offers improvement suggestions to guide students in optimizing their E-R diagrams and enhancing database design quality. Regarding SQL efficiency, instructors analyze whether teams' SQL statements contain performance bottlenecks—such as full table scans or improper index usage—and offer optimization suggestions like creating appropriate indexes or refining query structures. This dynamic feedback mechanism establishes an "evaluation-feedback-optimization" loop, enabling students to promptly identify issues, adjust learning strategies, and continuously enhance their learning outcomes and professional competencies.

5 Teaching Reflection and Improvement

5.1 Challenges Encountered During Reform. Teacher Resources and Class Time Constraints: Tiered task design and formative assessment require additional time investment. In project-based learning, students need ample group discussion to jointly analyze problems and explore solutions—a process demanding significant time. Teachers must provide personalized guidance tailored to each group's specific situation, address student inquiries, and assist with problem-solving, further requiring greater time and effort. However, limited class hours struggle to accommodate project-based learning's time demands, preventing certain teaching segments from being thoroughly developed and consequently impacting instructional outcomes.

Low Collaboration Motivation Among Some Students: A minority of Group A students relied excessively on peers while contributing minimally.

Addressing Foundational Gaps: Despite implementing tiered instruction and personalized guidance, a small number of students still faced significant challenges in abstract theoretical learning, such as paradigm identification. These students may struggle to grasp the principles and methods of paradigm determination due to weak mathematical foundations or insufficient logical thinking skills. For this group, additional learning resources such as online micro-lectures should be provided, enabling self-paced learning according to individual schedules. Concurrently, one-on-one tutoring should be offered to address learning challenges, ensuring all students master fundamental knowledge and skills.

5.2 Continuous Improvement Measures. To address the aforementioned issues, the following improvement measures are proposed:

Faculty Support: Establish a "Course Teaching Team" with divided responsibilities for "task design-assessment-question answering" to reduce individual workload; introduce an "automated grading system" to minimize time spent on mechanical grading.

Collaborative Management: Implement a "Contribution Metrics Table" (e.g., SQL code submissions, discussion participation) and mandate weekly "Team Contribution Reports" from each group. Instructors will conduct individual discussions with low-contributing students.

Develop an online micro-lecture resource repository to support personalized learning.

Resource Optimization: Invite industry mentors to participate in project defense evaluations and introduce a case repository featuring scenarios closer to real-world business contexts.

6 Conclusion

This study addresses teaching challenges in the Database Principles and Applications course at private universities by integrating CDIO-OBE principles. It constructs a teaching reform plan based on CDIO-OBE integration, effectively resolving issues such as disconnect between theory and practice, significant variations in student foundations, and monolithic evaluation methods through a reverse-designed goal system, tiered teaching content, differentiated implementation strategies, and a diversified assessment system. This demonstrates the applicability and effectiveness of CDIO-OBE in applied courses at private institutions while fostering teamwork and innovative thinking.

Future efforts will focus on deepening industry-education integration, exploring smart teaching assistance, continuously refining the reform plan, and conducting long-term tracking of graduate development to further validate and enhance this reform model.

References

[1] Zhou Rui, Bao Honghui. Research on Teaching Model for Research-Oriented Graduation Projects in Food Science and Engineering Based on CDIO-OBE [J]. Agricultural Products Processing, 2024, (22): 140-144. 1671-9646(X).2024.22.029. (In Chinese)

[2] Zu Yikang, Xu Miaojing, Xu Chunhui. Research on Teaching Model for Microcontroller Course Design Based on CDIO Concept [J]. Computer Education, 2019(9): 34-37. (In Chinese)

[3] Bai Erjing, Pang Shulan. Research on "OBE-CDIO" Teaching Reform for Applied Undergraduate Computer Majors in the Context of Accreditation Evaluation [J]. Fujian Light Industry and Textile, 2025(1):68-71. (In Chinese)

[4] Wang Junmei, Wu Jihong, Zheng Dongxia, et al. Exploration of Blended Teaching for Software Testing Courses Based on OBE and CDIO [J]. Software Engineering, 2019,22 (10):54-56. (In Chinese)

[5] Lu Dalin, Wu Bin. Tutorial on SQL Server 2008 Database Application and Development [M]. Beijing: Higher Education Press, 2020. (In Chinese)

[6] Zhang Ying, Feng Sang, Liu Yanwei, et al. Exploration of Theoretical Mechanics Teaching Reform Based on OBE-CDIO Integration for Vehicle Engineering [J]. Times Auto, 2025, (07): 74-76. (In Chinese)

[7] Luo Dongmei, He Shanshan, Sun Wenling, et al. Exploration and Practice of Embedded Systems Curriculum Reform Under CDIO-OBE Philosophy [J]. Computer Knowledge and Technology, 2025, 21(08): 161-164. 2025.0406. (In Chinese)

[8] Zhai Mingliang, Xie Jiucheng, Yu Liang, et al. Exploration of Reforming Python Programming Courses Using a Diverse Blended Approach Driven by OBE and CDIO Educational Philosophies [J]. Chinese Character Culture, 2024, (19): 181-183. 2024.19.064. (In Chinese)

[9] Wang Yan, Meng Yakun, Zhang Bin, et al. Research on Project-Driven Teaching Reform of Access Database Courses Based on CDIO [J]. Software, 2020, 41(05): 75-77+142. (In Chinese)

[10] Hu Taoying, Yu Juan. Analysis of Java Programming Course Reform Under CDIO-OBE Engineering Education Philosophy [J]. Computer Knowledge and Technology, 2024, 20(11): 137-139. 2024.0557. (In Chinese)

A Case Study on Integrating Professional Ethics into General Artificial Intelligence Education Empowered by AIGC

Rong Feng
Xi'an Technology and Business College, Xi'an, Shaanxi, China
283841921@qq.com

Abstract. To address the theory-practice gap in ethics instruction within artificial intelligence (AI) general education, this action research study developed a three-layer pedagogical model—"Problem Orientation, Implementation, and Value Shaping"—enhanced by AI technology. Centered on the core case study, "My Professional Assistant," and supported by complementary cases, the model was implemented over one semester. A mixed-methods approach, comprising pre- and post-test questionnaires, content analysis of student deliverables, and in-depth interviews, was employed for evaluation. Results demonstrated significant improvements in students' sensitivity to technological ethics ($p < 0.001$) and their knowledge of responsible AI practices ($p < 0.001$). Qualitative findings revealed that students transitioned from being mere tool users to responsible supervisors, integrating ethical considerations deeply into their professional practice. Furthermore, several student projects were adopted by external partners, generating impact beyond the classroom. This research provides an actionable framework and empirical support for the systematic cultivation of professional ethics literacy in general AI education.

Keywords: AI General Education; AIGC; Professional Ethics; Action Research; Mixed Methods; Value Internalization

1. Introduction

As intelligent technologies advance, they have become a transformative force empowering various disciplines. However, current general AI courses for non-computer science majors often remain limited to introducing tools and superficial applications, failing to adequately address the ethical challenges of AI integration into professional fields, such as issues of academic integrity, accountability, and data bias[1]. This model that prioritizes technical operation over ethical reflection may produce graduates who are technically proficient but ethically ungrounded, thereby amplifying the risks of AI misuse[2].

Cultivating students who can adhere to professional ethics and navigate the boundaries of human-computer collaboration has thus become an urgent mission for AI general education. However, theoretical instruction or case studies alone prove insufficient for fostering a genuine, internalized appreciation of ethical norms among students. This study therefore explores a central proposition: can professional ethics literacy be systematically fostered and internalized through well-designed technical practices[3]?

To this end, grounded in constructivism and action research, this study developed a three-layer teaching model: "Problem Orientation–Technical Implementation–Value Shaping."This study empirically examines the processes and outcomes of a teaching model designed to cultivate professional ethics in non-computer science majors. It employs a one-semester action research approach, centered on the core case "My Professional Assistant" and supported by cases such as "Dialect Guardian," utilizing mixed methods to collect and analyze multi-source data.

2. Theoretical Framework and Research Design

2.1 Theoretical Framework: The "Problem-Technology-Value" Three-Layer Model

Grounded in constructivism and outcome-based education (OBE), this study developed a three-layer teaching model termed the "Problem-Technology-Value" framework, which comprises "Problem

Orientation," "Technical Implementation," and "Value Shaping." The model's core principle is to transform ethics literacy from a topic of external instruction into an intrinsic part of hands-on technical practice[4].

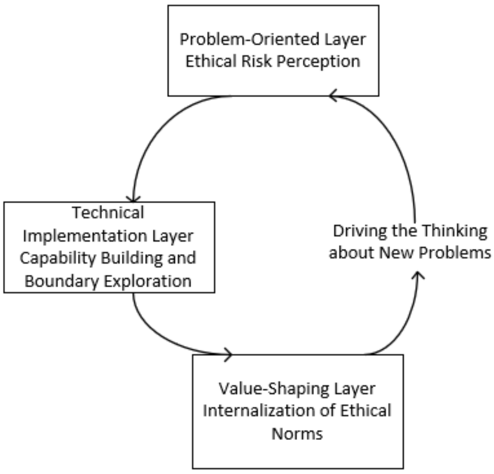

Figure 1. Theoretical framework of the three-layer "Problem-Technology-Value" teaching model

Problem Orientation Layer (Ethical Risk Perception): This layer begins by examining potential risks associated with AI applications across various professional disciplines. For example, in the core case "My Professional Assistant," the central question is "How to prevent intelligent assistants in law or medicine from overstepping their authority?" In the auxiliary case "Dialect Guardian," the issue explored is "Can AI-driven standardization inadvertently erode linguistic diversity?"

Technical Implementation Layer (Capability Building and Boundary Exploration): Students use accessible AI tools to build professional agents, train dialect models, or generate cultural content. This hands-on construction serves as a means for technical skill acquisition and, more importantly, for actively probing the capabilities and limitations of AI technologies through firsthand experience.

Value Shaping Layer (Ethical Norm Internalization): This stage guides students to integrate ethical reasoning with their technical practice, culminating in tangible deliverables such as a Professional Agent Ethics Charter or Ethical Guidelines for Dialect Data Collection. The process transforms external professional norms into a internalized, personal framework for technology ethics.

2.2 Research Methodology: Action Research Paradigm and Mixed Methods

This study employed an action research approach. Action research is characterized by iterative 'planning-acting-observing-reflecting' cycles, through which practitioners in real-world contexts simultaneously improve their practices and generate knowledge. This approach was ideally suited to the exploratory aims of our study, as it enabled the iterative refinement of teaching strategies in response to student feedback. This dynamic process allowed for an in-depth elucidation of the complex development of ethics literacy.

To comprehensively evaluate the teaching effectiveness, the study employed a mixed-methods approach to data collection and analysis, enabling methodological triangulation to enhance the validity and reliability of the findings. Data collection included:

1) Quantitative data: Pre- and post-test questionnaires were administered, incorporating an adapted Technology Ethics Sensitivity Scale and a self-developed Responsible AI Application Knowledge Test to evaluate student growth in ethical awareness and applied knowledge, respectively[5].

2) Qualitative data I: We performed a systematic content analysis of all final project reports, coding the texts to identify key themes and to evaluate the depth and sophistication of students' ethical reasoning.

3) Qualitative data II: Semi-structured interviews were held with 12 randomly selected students to gain deep insights into their evolving perspectives and learning experiences.

4) Outcome evidence: We collected documentation demonstrating the adoption of student projects by external organizations, serving as tangible validation of the curriculum's real-world impact.

3. Design and Implementation of Core and Auxiliary Case Clusters

This study adopted "My Professional Assistant" as its core case study, which tasked students with developing an ethically sound domain-specific assistant for their own majors. The instructional design followed a progressive 'Basic-Advanced-Challenge' framework, which unfolded over two complete action research cycles.

3.1 Implementation and Iteration of Teaching Actions

To systematically foster professional ethics literacy, this study developed a suite of pedagogical cases centered on a core project and supported by auxiliary cases. Organized around the central theme of professional ethics and responsibility boundaries, this case-based curriculum extends into distinct dimensions—including data ethics, cultural critique, and public ethics—thereby creating a comprehensive framework for developing ethics literacy. The structure of this framework and its corresponding learning objectives are detailed in Table 1.

Table 1 Structure and Learning Objectives of the Core and Auxiliary Cases

Case Name	Core Question	Layered Task Design	Core Literacy Dimension Focus
Core Case: My Professional Assistant	How to prevent ethical boundary-crossing and responsibility risks of professional agents?	Basic: Build professional knowledge base Advanced: Diagnose AI boundary-crossing behavior Challenge: Design human-machine collaborative compliance plan	Professional Ethics & Responsibility Boundaries
Auxiliary Case 1: Dialect Guardian	Does technological efficiency priority sacrifice cultural diversity and fairness?	Basic: Collect and annotate dialect data Advanced: Train and optimize dialect model Challenge: Formulate service and inheritance plan	Data Ethics & Cultural Responsibility
Case Name	Core Question	Layered Task Design	Core Literacy Dimension Focus
Auxiliary Case 2: My Chang'an Poetic Charm	How to address AI's misinterpretation and stereotypical output in cultural creation?	Basic: Generate cultural theme content Advanced: Diagnose and correct cultural misinterpretations Challenge: Design cultural promotion plan	Critical Thinking & Cultural Awareness
Auxiliary Case 3: My City Brain	Do smart technology solutions consider inclusivity and public value?	Basic: Identify urban governance problems Advanced: Diagnose resource misallocation and waste Challenge: Design inclusive optimization plan	Public Ethics & Social Responsibility

As outlined in Table 1, all cases adhere to a "Basic-Advanced-Challenge" task structure, mirroring the "Problem-Technology-Value" pedagogical model to ensure a progressive journey from technical practice to value internalization. The core case, "My Professional Assistant," acts as a primary vehicle for exploring responsibility boundaries within professional contexts; the auxiliary cases extend these explorations into data ethics, cultural critique, and public governance, collectively demonstrating the transferability of the ethical literacy developed through the core case.

This study adopted "My Professional Assistant" as the core case and undertook a one-semester action research study comprising two cycles. The case tasked students with developing an ethically sound and professionally appropriate intelligent assistant for their respective majors.

First Cycle (Basic-Advanced Layer): Students employed RAG technology to construct a professional knowledge base and to identify situations where the agent might overstep its authority. Observation revealed that some students initially held a superficial understanding of what constituted an ethical boundary. In response, the second cycle incorporated a 'Professional Ethics Clause Comparison' task,

requiring students to evaluate the agent's outputs against explicit industry regulations. This intervention significantly sharpened their critical analysis.
Second Cycle (Challenge Layer): Students were tasked with designing a "Human-Machine Collaborative Compliance Plan." However, analysis of initial submissions revealed that their plans showed limited consideration of broader societal values. To address this, auxiliary cases such as "My City Brain" were introduced to provide cross-contextual insights, which led them to integrate ethical considerations like inclusivity and equity into their revised plans.
The auxiliary cases were strategically deployed throughout the curriculum to reinforce and extend the ethical competencies developed in the core case, thereby constituting a dynamic and iterative instructional system.

4. Analysis of Teaching Effectiveness

To evaluate the model's effectiveness in fostering professional ethics literacy, data were analyzed using a mixed-methods approach.

4.1 Quantitative Evidence: Significant Improvement in Ethical Awareness and Knowledge

Pre- and post-test questionnaires administered to the 58 students who completed the course revealed statistically significant improvements in both "Technology Ethics Sensitivity" and "Responsible AIGC Application Knowledge" (see Table 2).

Table 2 Pre- and Post-Test Comparisons of Student Ethical Literacy (N = 58)

Measurement Dimension	Pre-test (M±SD)	Post-test (M±SD)	t-value	p-value
Technology Ethics Sensitivity	3.15 ± 0.62	4.32 ± 0.48	9.45	< 0.001
AIGC Responsible Application Knowledge	66.8 ± 12.4	86.5 ± 9.7	8.71	< 0.001

The data indicate that the teaching intervention significantly enhanced students' ethical awareness and strengthened their knowledge of key AIGC responsibility and boundary issues.

4.2 Qualitative Evidence I: Deep Ethical Thinking in Project Reports

Thematic Analysis: Depth of Ethical Internalization：

1) Semantic Shift in Student Discourse
Words such as "responsibility," "boundary," "bias," and "fairness" appeared 350% more frequently in final reports than in early-course discussions, indicating a pronounced shift in students' ethical focus.
2) Contextualized Boundary Analysis
All reports identified at least one type of AI boundary violation, with over 70% providing professionally contextualized risk and consequence analyses. For instance, law students examined liability for AI-generated legal documents.
3) From Compliance to Value Propositions
Eighty-five percent of compliance designs transcended technical compliance to proactively embed societal values such as supporting vulnerable groups, preserving cultural heritage, and ensuring public safety.

4.3 Qualitative Evidence II: Role Transformation and Internalization Process Revealed in Interviews

Inductive analysis of the interview data revealed key trajectories in the students' role transformation:
Theme 1: "Awakening from Tool Awareness to Responsibility Awareness." One student 坦言: "In the first half of the course, I only cared about how to make the agent answer more accurately and quickly. But now, I first consider where the bottom line of my profession lies and whether the agent's advice might mislead others."
Theme 2: "Failure is the Most Profound Teacher." Several students mentioned that the "dangerous errors" (e.g., a medical agent recommending an overdose) that occurred while debugging the agent

gave them a great psychological impact. It was these "failure" experiences that made them deeply appreciate the necessity of "human-in-the-loop."

Theme 3: "The Sense of Responsibility from Being Recognized." A student whose "Campus Language Accessibility Service Construction Plan" was adopted by the logistics department said: "When your idea can actually become a service at the university, you truly feel the heavy social responsibility behind the technology."

4.4 Social Value Spillover of Learning Outcomes

Multiple student-generated outcomes from this course were successfully translated from the classroom into real-world practice. For example, the Ethical Guidelines for Legal AI Agents was adopted by the Law School as part of its freshman orientation materials, while the Campus Language Accessibility Service Proposal received formal endorsement from the university's logistics department and entered the feasibility assessment phase. This "learn–apply–create impact" cycle significantly strengthened students' sense of achievement and reinforced the notion of ethics literacy as a tangible responsibility in professional contexts.

5. Discussion and Implications

This study, through a completed action research cycle, offers the following insights for ethics instruction in general AI education:

Ethics Education Must Be Embedded in the "Capillaries" of Technical Practice: Our implementation demonstrates that ethics is not an abstract set of rules superimposed on technology, but an embodied understanding that emerges organically through each step of building, debugging, failing, and refining AI agents. This "learning ethics by doing" approach fosters the integration of knowledge and practice.

"Responsibility Boundary" is a Trainable, Internalizable Core Literacy: Through repeated diagnosis and regulation of boundary-crossing behaviors in the core case, students developed a "boundary awareness" that transferred effectively to other technological contexts. Qualitative evidence indicates that this internalized competency manifests as a proactive and anticipatory habit of ethical reflection.

Action Research and Mixed Methods are Effective Paths for Exploring Value-Shaping Courses: For courses aiming to shape complex internal constructs such as values and attitudes, quantitative approaches alone are insufficient. This study shows that mixed-methods action research—integrating quantitative trend analysis with in-depth qualitative insight—can deliver rigorous and comprehensive evidence to support such educational innovations.

6. Conclusion and Outlook

Guided by the action research methodology, this study implemented a "Problem–Technology–Value" teaching model and employed mixed methods to evaluate its effectiveness. The results empirically demonstrate that professional ethics literacy can be effectively developed, observed, and internalized in general AI education. By engaging in technical practices that involved defining boundaries and assigning responsibility, students transformed external ethical norms into stable personal values and behavioral guidelines.

Future research will advance along two trajectories: first, to develop AI-supported instructional systems capable of performing preliminary ethical analysis of student project reports, thereby offering educators real-time feedback; and second, to conduct multi-institutional longitudinal studies assessing the long-term efficacy and generalizability of the training model. These efforts aim to further advance educational strategies for fostering responsible digital citizens in the age of artificial intelligence.

Acknowledgements

This work was supported by the 2025 Xi'an Technology and Business College Teaching Reform Project "AIGC-Empowered Reform and Practice of Case-Based Teaching in Artificial Intelligence Courses" (Grant No. 25KGZX04).

References

[1] He, S., Wu, F., & Zheng, J. (2023). Digital transformation enables teaching reform practice in digital graphic information processing technology specialty group. Printing and Digital Media Technology Research, (4), 41-46.

[2] Zhou, H., & Tan, Q. (2025). Practical exploration of teaching model innovation in design driven by industry-education integration and enabled by AIGC technology. Printing and Digital Media Technology Research, (S1), 25-29.

[3] Bao, P., Xing, W., & Lu, W. (2021). Research on teaching model innovation of artificial intelligence practice courses under the background of emerging engineering education. Computer Education, (6), 105-109.

[4] Fang, B., & Hu, R. (2019). Artificial intelligence schools in Chinese universities: Current situation, problems and development direction. China Science and Technology Industry, (9), 70-75.

[5] Sun, L., Su, D., & Tang, B. (2022). Research on optimization strategy of talent training mode for emerging engineering education in local universities: Based on the outcome-based perspective in engineering education certification. Journal of Heilongjiang Institute of Teacher Development, (9), 7-11.

Undergraduate Graduation Design Practice in Computer Science and Technology

Yuxuan Dong[1, a *]

[1]Xi'an Technological University, Xi'an, China

[a]2174207343@qq.com

Abstract. This paper takes the undergraduate graduation design of the Computer Science and Technology major as the research object, and fully records the entire process of the project "Design and Implementation of a Rapid Goods Recognition System for Cabinets Based on Deep Learning" from initiation to implementation. It focuses on analyzing the key actions and core gains in the stages of topic selection and decision-making, technical preparation, engineering practice, and reflection and summary. In the topic selection stage, it breaks through the limitations of traditional development directions and finally selects the uncontacted deep learning field among small program development, conventional system design, and deep learning applications. In the technical preparation stage, it builds a theoretical framework for target detection and system development based on pre-graduate study during the winter vacation and literature review. In the engineering practice stage, aiming at the problem of insufficient computing power of personal equipment, it systematically explores server rental and selection schemes, and overcomes a series of technical difficulties in environment setup, code reproduction, and function integration. Finally, a cabinet goods recognition system with image detection, video recognition, and result export functions is completed. The research confirms that this graduation design not only realizes the cross-field integration of professional knowledge but also cultivates the abilities of independent learning, problem diagnosis, and engineering implementation. It provides a reusable graduation design practice paradigm for computer major students and helps them efficiently complete the core tasks in the final stage of their studies.

Keywords: Undergraduate Graduation; Deep Learning; Cabinet Goods Recognition; Server Configuration; YOLOv7 Model

1. Introduction

As the core practical link of the four-year academic studies of the Computer Science and Technology major, undergraduate graduation design is not only a comprehensive test of professional knowledge such as programming languages, algorithm principles, and system architecture but also an important link connecting classroom theory and industrial applications [1]. Different from disciplines that focus on physical manufacturing, such as mechanical engineering, the graduation design of the computer major has significant technical iteration characteristics — from traditional management system development to the application of artificial intelligence, the difference in topic selection directions directly determines the depth of practice and the dimension of ability growth.

At the beginning of the graduation design, I faced the common "technical route selection dilemma" of computer major students: on the one hand, I hoped to transform the already mastered programming languages such as Python and Java into practical applications; on the other hand, I was worried that the topic selection would fall into the inefficient misunderstanding of "repeated development". Initially, three alternative directions were formulated: the first is lightweight WeChat mini-program development, whose technology stack (Vue, WeChat Developer Tools) has been practiced in the course design, with low development difficulty but limited innovation value; the second is the traditional B/S architecture management system, which adopts the SpringBoot MySQL architecture, with a mature development process but difficult to reflect technical breakthroughs; the third is the deep learning-driven cabinet goods recognition application, which involves uncontacted fields such as computer vision and neural networks but is in line with the actual demand of "unmanned inventory" in the intelligent retail industry. This entanglement is essentially a game between the "safe choice in the comfort zone" and the "growth choice in the challenge zone" — the former can ensure the timely completion of tasks, while the latter may bring unexpected ability improvement [2].

Finally, the deep learning direction is determined. The core reason lies in the practical value and technical integration of the cabinet goods recognition scenario: this scenario not only requires model training capabilities (the core of deep learning) but also requires system development capabilities (interface and function implementation), and also involves engineering issues such as data processing and hardware adaptation, which can comprehensively exercise the comprehensive quality of computer majors. Just as the law of "technology iteration precedes application implementation" in the computer field, actively embracing the unmastered technical field can make the graduation design an opportunity to break through the boundary of one's own abilities, rather than a simple application of knowledge [3].

2 Topic Selection Decision-Making and Preliminary Preparation

2.1 Topic Selection Balance

The Decision-Making Process of Breaking Through the Technical Comfort Zone. The core contradiction in the topic selection of the computer major's graduation design lies in the balance between "technical familiarity" and "practical innovation".

The three alternative directions I initially had each had obvious advantages and disadvantages:

1) Mini-program development: The advantage is that it is based on the WeChat ecosystem, with low difficulty in user interaction design, and relevant experience has been accumulated in the "campus information query" course design; the disadvantage is that the functions are mostly concentrated on data display and form submission, which is difficult to reflect the technical depth of the computer major, and the market has a saturation of similar applications, lacking practical application value.

2) Traditional management system: The advantage is that the development process is standardized, the link from demand analysis to deployment and launch is clear, and the function implementation can be completed quickly; the disadvantage is that the technology stack is relatively traditional, which cannot touch the current hot directions in the computer field (such as artificial intelligence and big data), and the graduation design results have limited support for future career development.

3) Deep learning-based cabinet recognition: The advantage is that it is in line with the pain points of the intelligent retail industry (low efficiency and high error rate of manual inventory), can integrate multi-dimensional technologies such as target detection models, data annotation, and system development, and has public datasets (such as VOC2007) and mature models (YOLO series) to reduce the entry threshold; the disadvantage is that it is necessary to learn knowledge such as neural network principles and model training from scratch, and the hardware computing power requirement may exceed the carrying capacity of personal equipment.

During the topic selection process, three in-depth communications with the supervisor played a key role in promotion. During the first communication, I put forward the concern of "lack of deep learning foundation", and the supervisor pointed out: "The core competitiveness of the computer major lies in the ability to quickly learn new technologies, and the graduation design should be a carrier to exercise this ability, rather than a simple application of knowledge"; the second communication focused on the feasibility of the scenario, and the supervisor suggested giving priority to the cabinet recognition direction with "easy access to data and open-source foundation for models" to avoid falling into the complex trap of "algorithm innovation"; the third communication determined the specific technical route, and clarified that YOLOv7 is used as the basic model, the core recognition function is realized first, and then the visual interface is expanded to ensure the practical logic of "first implementation, then optimization" [4].

Finally, the topic is determined as "Design and Implementation of a Rapid Goods Recognition System for Cabinets Based on Deep Learning". This choice not only avoids the inefficient dilemma of "repeated development" but also endows the graduation design with innovation through "cross-field technology integration". More importantly, it forces me to jump out of the limitation of "only being able to write basic code" and face the dual challenges of theory and engineering in the deep learning field.

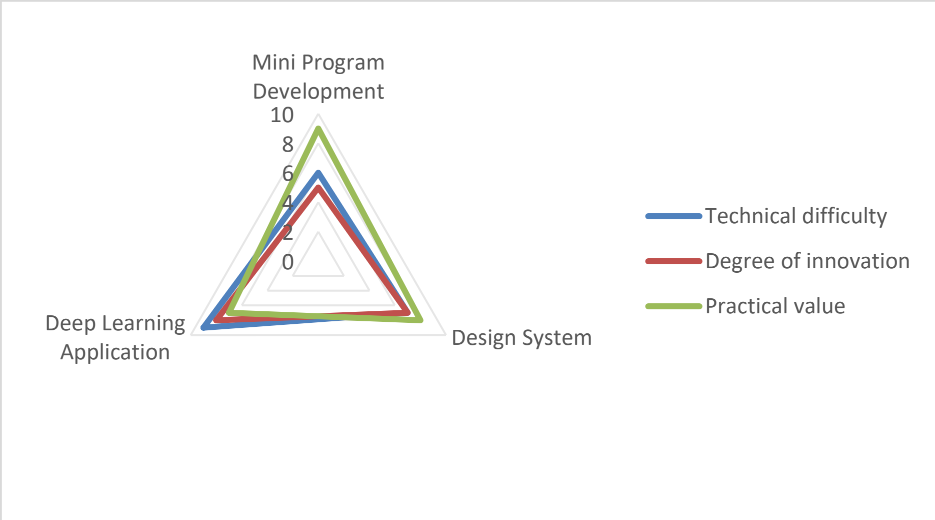

Figure 1. Trade-off Diagram for Topic Selection Decision-making

2.2 Pre-Graduation Study During Winter Vacation: Building a Systematic Technical Foundation

To reduce the technical resistance after the start of the graduation design, I carried out pre-graduate study for 40 days during the winter vacation, adopting a three-dimensional learning model of "video practice + literature review + tool rehearsal" to ensure that I could quickly enter the practical stage after the start of the new semester:

1) Theoretical introduction: Through the Bilibili series course "Deep Learning and Computer Vision Practice", I systematically learned the basic structure of neural networks, the principle of convolution operations, and the core process of target detection (such as the "end-to-end detection" mechanism of the YOLO model). I focused on understanding the complete link of "feature extraction → bounding box prediction → category judgment", compared the advantages and disadvantages of Faster R-CNN, SSD, and YOLO series models, and clarified the balanced advantages of YOLOv7 in "real-time performance" and "accuracy" — this is crucial for the "real-time inventory" demand of cabinet goods recognition [5].

2) Literature research: Using the school library database (CNKI, IEEE Xplore), I searched for literatures related to "cabinet goods recognition", "small target detection optimization", and "YOLO model improvement", and sorted out three core technical pain points in this field: complex background (shelf texture) interfering with recognition, missing detection of small-sized goods (such as chewing gum and toothpicks), and low accuracy in occluded scenarios (goods stacking), which provided directions for subsequent model optimization; at the same time, I referred to 10 undergraduate graduation design literatures to learn the writing logic of "theoretical derivation → experimental verification → system implementation" to avoid logical confusion in the later thesis writing.

3) Tool preview: I installed basic tools such as Python 3.8, PyCharm, and Anaconda on the personal computer in advance, practiced operations such as image reading, cropping, and format conversion of OpenCV through simple cases, was familiar with the environment creation and package management commands of Anaconda (such as conda create and conda install), and tried to use the LabelImg tool to complete a small amount of image annotation to avoid wasting time on basic tool operations in the later stage.

The greatest gain of the pre-graduate study is the establishment of a "technical knowledge graph": it is clear that the cabinet goods recognition system needs to be decomposed into three modules: "data collection and preprocessing → model training and optimization → system development and visualization", and each module is further subdivided into specific technical points (for example, data preprocessing needs to include annotation, enhancement, and division; model training needs to involve parameter tuning and computing power adaptation). This structured cognition avoids the inefficient problem of "blind trial and error" in the subsequent practice [5].

3 Technical Breakthrough

3.1 Computing Power Bottleneck: Hardware Limitations of Personal Equipment

In the stage of preparing for model training, the typical obstacle of deep learning projects in the computer major was first encountered — insufficient computing power of personal equipment. The laptop I used was equipped with an NVIDIA MX350 graphics card with only 2GB of video memory, while the minimum video memory requirement for YOLOv7 model training is 4GB (lightweight YOLOv7-tiny), and the complete model requires more than 8GB of video memory. Initially, an attempt was made to reduce the model complexity: the lightweight YOLOv7-tiny was used, and the batch_size (number of batch training samples) was reduced to 4. As a result, it was found that the recognition accuracy of the model for small-sized goods dropped by 32%, which could not meet the core demand of "small packaged goods recognition" in the cabinet scenario at all [7].

This problem is not an individual case but a common dilemma for computer major students to carry out deep learning projects: personal equipment is limited by cost and portability and cannot meet the high computing power requirements of model training. If it cannot be solved, the entire graduation design will fall into the deadlock of "theoretical design cannot be implemented and verified" — the core value of deep learning projects lies in the "actual data verification of model effects", rather than just staying in code writing and theoretical derivation [8].

3.2 Server Rental: A Complete Exploration from Selection to Practice

Aiming at the problem of insufficient computing power, I systematically investigated the mainstream server rental platforms and configuration schemes, and formed a scientific selection strategy by comparing the three dimensions of "configuration flexibility", "price cost", and "ease of use":

1) Selection: General cloud servers such as Alibaba Cloud and Tencent Cloud were prioritized for exclusion because such platforms require manual configuration of deep learning environments (such as installing CUDA and PyTorch), which are not user-friendly for beginners; finally, deep learning-specific platforms such as AutoDL and Jilian AI Cloud were selected. These platforms are pre-installed with mainstream frameworks and dependency libraries and support "hourly billing", which is suitable for the limited budget of students (the average daily cost is controlled at 15-20 yuan).

2) Configuration selection: The core focuses on four parameters: GPU model, video memory size, CPU core number, and memory capacity. Initially, a Tesla T4 graphics card with 2GB video memory was tested, and the "CUDA out of memory" (video memory overflow) error occurred frequently during training; after switching to an RTX 3060 graphics card with 8GB video memory, it could support batch training with batch_size=16, and 100 rounds of training only took 8 hours (4 times faster than the T4 graphics card); at the same time, it was equipped with a 4-core CPU and 16GB memory to ensure that data loading (such as reading large datasets) and model reasoning do not get stuck, avoiding the resource waste of "GPU idle waiting for CPU".

3) Remote operation: Code writing and training monitoring were completed through the Jupyter Notebook provided by the platform, and SSH tools (such as PuTTY) were used to realize file transmission between the local and the server — for example, the locally annotated dataset was uploaded to the server through the SFTP protocol to avoid data loss caused by network fluctuations; during the training process, the model weight file (checkpoint) was saved every 10 rounds to prevent the loss of training results due to server expiration or connection interruption.

In the practice of server rental, two major problems were encountered: first, the "RTX 3090 with 16GB video memory" configuration was selected initially, with a daily cost of 50 yuan, which far exceeded the budget. Later, through the method of "first testing the feasibility with low configuration", it was determined that the RTX 3060 with 8GB video memory was the most cost-effective solution; second, the training was terminated due to remote connection interruption, which was later solved by adding "automatic resumption logic" in the code (reading the latest weight file to continue training).

This experience made me deeply realize that computer engineering practice is not only "code implementation" but also includes hidden capabilities such as hardware resource planning and cost control — these capabilities are often ignored in classroom learning but are crucial in the actual project implementation [9].

4 System Development

4.1 Environment Setup: The Basic Engineering of Deep Learning Projects

Environment setup is the first technical threshold of deep learning projects. The core challenge is to ensure the compatibility of Python version, framework version, and dependency library version — any version mismatch may cause the model to fail to call the GPU or have abnormal functions. Initially, the installation command was directly copied by referring to the open-source project blog. As a result, due to the incompatibility between the PyTorch version and the CUDA version, the problem of "torch.cuda.is_available() returning False" occurred. Later, through the method of "step-by-step verification and layer-by-layer troubleshooting", a stable environment was set up in 3 days:

1) Version matching confirmation: The server was pre-installed with CUDA 11.6. According to the "CUDA version - framework version" correspondence table on the PyTorch official website, the versions of "torch==1.13.1+cu116" and "torchvision==0.14.1+cu116" were selected to avoid GPU call failure caused by incompatibility between the framework and the hardware interface.

2) Batch management of dependency libraries: A requirements.txt file was created to uniformly manage core dependency libraries, including opencv-python (image processing, version 4.6.0.66), pandas (data statistics, version 1.5.3), pyqt5 (interface development, version 1.5.7), and ultralytics (YOLOv7 implementation, version 8.0.12). The command "pip install -r requirements.txt" was used for batch installation to reduce the risk of version conflicts caused by manual installation.

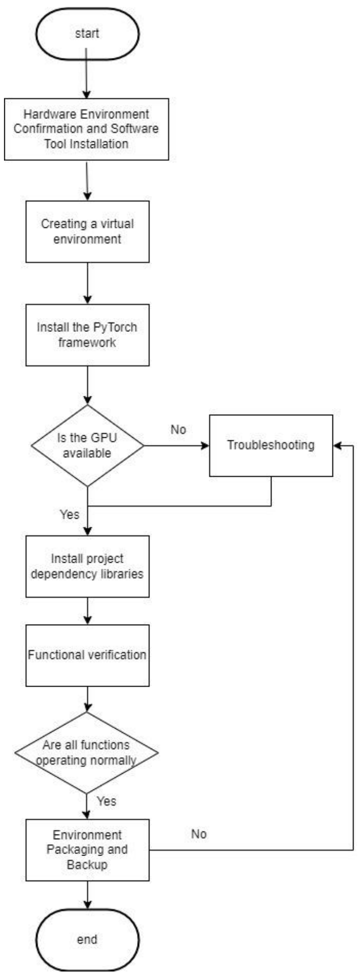

Figure 2. Environment Setup Flowchart

3) Functional verification test: Three sections of test code were written to verify the stability of the environment: the first was GPU availability test (printing torch.cuda.device_count() to confirm device recognition); the second was image processing test (reading and displaying a cabinet goods image with OpenCV); the third was model loading test (loading the pre-trained YOLOv7 weight file to infer the test image). Formal development was carried out only after ensuring that there were no abnormalities in each link.

During the environment setup process, problems such as "OpenCV failing to read images with Chinese paths" and "PyQt5 interface displaying garbled characters" were also solved: the former replaced cv2.imread() with the method of "np.fromfile() reading binary data + cv2.imdecode() decoding"; the latter solved the problem by setting the Chinese font (such as "Microsoft YaHei") of "QtGui.QFont". These seemingly trivial problems are actually important practices of the "compatibility thinking" in computer system development — the stability of engineering projects often depends on the ability to handle detailed problems [10].

4.2 Code Reproduction: The Technical Transition from Imitation to Understanding

Code reproduction is the core link of deep learning projects. The purpose is to lay a foundation for subsequent model optimization and function integration on the basis of understanding the logic of open-source projects. I selected the open-source project "YOLOv7 Commodity Detection" with more than 5k stars on GitHub as the reproduction object, and completed the reproduction and problem troubleshooting in 2 weeks according to the process of "data preparation → model training → inference verification":

1) Data preparation: A hybrid dataset of "public dataset + self-built dataset" was constructed: the public dataset selected commodity category images in VOC2007 (a total of 5000 images), and the self-built dataset was obtained by offline shooting of supermarket cabinets (including 8 categories of goods such as food and daily necessities, a total of 3000 images); the LabelImg tool was used for annotation in PASCAL VOC format, and the annotation content included the bounding box coordinates and category names of the goods (such as "beverages", "snacks", "daily necessities"); the dataset was divided into a training set (7200 images) and a verification set (800 images) in a ratio of 9:1, and data enhancement technologies such as rotation (0°, 90°, 180°), horizontal flipping, and adding Gaussian noise were used to improve the generalization ability of the model to different scenarios.

2) Model training: Training parameters were configured based on the server's RTX 3060 graphics card: total rounds (epochs) of 100, batch size (batch_size) of 16, initial learning rate of 0.01, and the optimizer selected SGD (with momentum of 0.9); during the training process, TensorBoard was used to monitor key indicators in real time: training losses (obj loss, cls loss, box loss) and verification indicators (precision, recall, mAP@0.5) to ensure that the loss curve decreased steadily and the evaluation indicators continued to improve.

3) Inference verification: 800 images in the verification set were used to test the model effect, focusing on analyzing two types of problems: the missing detection rate of small-sized goods (such as chewing gum and toothpicks) and the false detection rate of occluded goods (such as stacked snack bags); a visualization tool (such as Matplotlib) was used to draw the detection results, and the commodity category, confidence, and bounding box were annotated to intuitively judge the performance shortcomings of the model.

The solution to two typical problems in the code reproduction process gave me a deeper understanding of deep learning engineering practice: first, "data annotation format errors" caused the model to fail to read training data. By comparing the standard VOC format XML files, problems such as "coordinate values exceeding the image resolution range" and "inconsistent case of category names" were corrected; second, "weight file damage after training interruption" was solved by adding the "checkpoint automatic saving" function in the training code (saving weights every 10 rounds and saving the optimizer state at the same time), ensuring that training could be resumed from the nearest node after interruption and avoiding the waste of computing power in the early stage [11]. It is worth noting that code reproduction is not "copying code", but understanding the model operation logic through the cycle of "modifying parameters → observing results → analyzing reasons" — for example, observing the fluctuation of the loss curve after adjusting the learning rate to master the influence law of "learning rate decay" on model convergence.

4.3 System Development: The Implementation Practice from Model to Application

After completing the model training, the system development stage was entered. The core goal was to transform "model inference capability" into "user-operable application functions" and finally realize the visualization and ease of use of "cabinet goods recognition". The system adopted a C/S architecture and built a desktop application based on PyQt5. The core functions included local image detection, real-time video detection, and recognition result export. The development process was divided into three modules:

1) Interface module: A GUI layout of "three areas and two columns" was designed — the left side was the "function selection area" (buttons for image upload and video access), the middle was the "result display area" (real-time display of detected images/videos, with commodity bounding boxes and category labels superimposed), and the right side was the "data list area" (displaying the name, confidence, and quantity statistics of recognized goods); the Qt Designer tool was used to complete the interface visualization design, and then PyUIC was used to convert the .ui file into Python code to ensure the interface was beautiful and the operation was intuitive.

2) Detection module: The YOLOv7 model was encapsulated into an independent "detection interface". The input was the image path or video stream frame, and the output was the commodity category, confidence, and bounding box coordinates; the detection logic was optimized for the cabinet scenario: an "image preprocessing" step (such as brightness equalization and denoising) was added to solve the problem of low recognition accuracy in strong/weak light environments; multi-threading technology was used to separate "image collection" and "model inference" to avoid interface lag caused by inference time consumption — for example, during video detection, the main thread was responsible for reading camera frame data, and the sub-thread was responsible for model inference. The results were transmitted through the signal-slot mechanism to ensure that the frame rate was stable above 20 FPS.

3) Export module: The recognition results could be exported in two formats: one was an Excel table (including commodity name, confidence, and detection time), which was convenient for users to count cabinet inventory; the other was annotated images/videos (superimposing detection boxes and category labels on the original images/videos), which was convenient for intuitively verifying the recognition effect; the export function was implemented through the pandas library (Excel generation) and OpenCV library (video writing), and at the same time, an "export progress bar" and "abnormal prompt" (such as non-existent file path) were added to improve the user experience.

The most challenging part of the system development stage was "multi-technical stack integration" — it was necessary to take into account model performance, interface response speed, and function stability at the same time. For example, initially, model inference and interface drawing were placed in the same thread, resulting in a video detection frame rate of only 5 FPS, which could not meet the real-time requirement; later, an independent "inference thread" and "interface refresh thread" were created, and thread isolation was realized using Qt's QThread class. At the same time, the signal-slot mechanism was used to safely transmit data, and the frame rate was increased to 22 FPS, which fully met the real-time inventory requirement of the cabinet scenario [12]. In addition, aiming at the problem of "false detection caused by dense stacking of goods", the model post-processing logic (such as adjusting the threshold of non-maximum suppression NMS) was optimized, and the false detection rate was reduced from 18% to 8%, further improving the practicality of the system.

5 Summary and Reflection on Graduation Design

5.1 Ability Improvement: Comprehensive Growth Beyond Technology Itself

Looking back on the entire graduation design process, the gains went far beyond "completing a operable system". More importantly, it cultivated the core literacy and engineering thinking of the computer major, which was specifically reflected in three dimensions:

1) Independent learning ability: From "zero foundation" to mastering cross-field technologies such as YOLOv7 model training, server configuration, and PyQt5 development, an efficient learning path of "official documentation first → technical blog supplement → community question verification" was formed. For example, when learning remote server connection, first refer to the AutoDL official help documentation, then refer to the practical cases of CSDN blog, and finally solve the "SSH connection timeout" problem on Stack Overflow. This independent learning mode is far better than passive acceptance of knowledge.

2) Problem-solving ability: Facing a series of problems such as "insufficient computing power", "environment conflict", and "code bug", a standardized process of "phenomenon description → cause location → scheme verification → summary and precipitation" was gradually established. For example, when solving the problem of "model training loss not decreasing", first observe the trend of the loss curve through TensorBoard, then check three possible reasons: dataset annotation format, learning rate parameters, and model weight initialization, and finally locate the problem as "annotation coordinate error causing the model to fail to learn effective features". This structured problem diagnosis ability is the core of computer engineering practice.

3) Engineering thinking ability: The focus shifted from "only paying attention to code correctness" to "taking into account functions, performance, cost, and user experience". For example, when selecting a server, the balance between "configuration level" and "budget limit" was considered; when developing the system, "interface ease of use" and "detection real-time performance" were considered; when optimizing the model, the balance between "accuracy" and "inference speed" was considered — this kind of thinking of "finding the optimal solution under constraints" is the key transformation of the computer major from "student" to "engineer" [13].

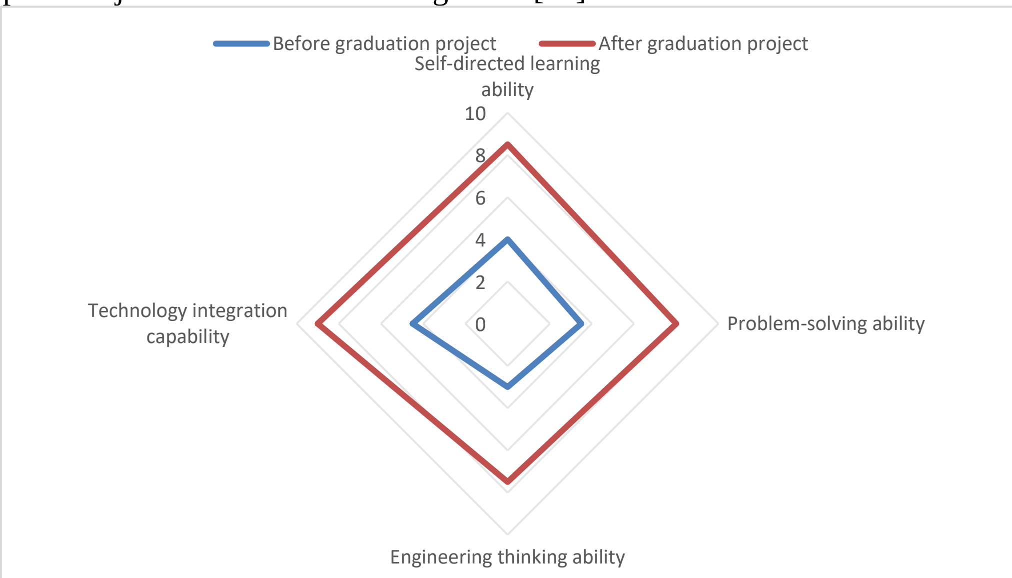

Figure 3. Competency Enhancement Radar Chart

5.2 Problems and Deficiencies

Although the system has realized the core functions, there are still three obvious deficiencies that need to be improved in subsequent optimizations:

First, the recognition accuracy of small-sized goods needs to be improved. The missing detection rate of goods with a size of less than 32×32 pixels (such as toothpicks and small packaged candies) is about 15%. The core reason is that the shallow feature extraction ability of the model is insufficient, and the underlying structure of the Feature Pyramid Network (FPN) for small targets has not been optimized;

Second, the system compatibility is limited. At present, it only supports Windows 10/11 systems and is not adapted to Linux or embedded devices (such as Raspberry Pi), which cannot meet the actual demand of "edge deployment" in the cabinet scenario, and the interface adaptation problem of screens with different resolutions is not considered;

Third, there is a lack of user feedback iteration. During the system development process, no actual cabinet operators were invited to test, which may lead to problems such as "the operation process not conforming to actual usage habits" (such as single export format and no support for batch image detection), affecting the practical application value. These deficiencies are essentially caused by "insufficient engineering practice experience" and "insufficient scenario awareness" — for example, the actual operation process of cabinet operators was not fully investigated, and functions were designed only based on their own understanding, leading to deviations between the system and actual needs.

5.3 Suggestions for Computer Major Graduation Design

Based on my own practice, five suggestions are put forward for the graduation design of junior students majoring in computer science and technology to help them complete the project efficiently:

1) "Dare to jump and try" in the topic selection stage: Priority should be given to directions with "technical challenges but open-source foundations" to avoid no improvement in ability due to "too familiar technology"; if a cross-field direction such as deep learning is selected, it is necessary to confirm that there are public datasets and mature models to reduce the entry threshold.

2) "Pre-start" in the preliminary preparation: Use winter vacation or spare time to learn core technologies in advance (such as model principles and development frameworks) instead of starting after the topic is determined; it is recommended to make a "technical knowledge graph" to clarify the learning objectives and outputs of each stage and avoid blind learning.

3) "Advance planning" for hardware problems: If the project involves deep learning, test the computing power of personal equipment as soon as possible. If it is insufficient, investigate server rental schemes in advance, compare the prices and ease of use of different platforms, and avoid delaying the project progress due to computing power problems.

4) "Iterative advancement" in the development process: Split the graduation design into small stages of "environment setup → core functions → optimization and iteration", and set clear delivery standards for each stage (such as successful model inference after environment setup), to avoid irreparable problems due to "last-minute rush"; at the same time, adhere to "developing while recording", and organize key problems and solutions into documents to facilitate later thesis writing.

5) "Proactive communication and collaboration" in the communication stage: Communicate with the supervisor regularly on the progress, especially when encountering technical bottlenecks, it is necessary to consult with "preliminary ideas and specific problems" instead of passively waiting for solutions; at the same time, a "technical exchange group" can be established with classmates to share environment configuration and code debugging experience and improve the efficiency of problem-solving [14].

6 Conclusion

The graduation design of "Design and Implementation of a Rapid Goods Recognition System for Cabinets Based on Deep Learning" is essentially a "cross-field integration practice" of professional knowledge in the Computer Science and Technology major — from technical confusion during topic selection, to exploration and attempt of server configuration, to detailed breakthroughs in environment setup, and to function integration in system implementation. Each link is a reconstruction and improvement of "computer professional ability". This process made me deeply realize that the core competitiveness of the computer major is not only "the ability to write code" but also "the ability to learn new technologies", "the ability to solve practical problems", and "the ability to implement innovative applications".

For computer major students, the value of graduation design does not lie in "creating a perfect system" but in "breaking through their own limitations in the process" [6].From "being afraid of contacting deep learning" to "independently completing model training and system development", this leap not only brings growth at the technical level but also maturity in mentality — understanding that in the computer technology industry with rapid iteration, "continuous learning" and "daring to challenge" are the long-term competitiveness. Just like the implementation process of the cabinet goods recognition system from "theoretical model" to "practical application", the undergraduate graduation design also completes the role transition from "classroom student" to "quasi-engineering and technical personnel", laying a solid practical foundation for subsequent career development.

References

[6] Brown A R ,Adams C J ,Ferner C , et al.Teaching parallel design patterns to undergraduates in computer science[C]//St. Olaf College, Northfield, MN, USA;;Calvin College, Grand Rapids, MI, USA;;University of North Carolina Wilmington, Wilmington, NC, USA;;Macalester College, Saint Paul, MN, USA;;University of North Carolina at Charlotte, Charlotte, NC, USA,2014:

[7] Prajish P ,Sridhar I .VeriSIM: A model-based learning pedagogy for fostering software design evaluation skills in computer science undergraduates[J].Research and Practice in Technology Enhanced Learning,2022,17(1):DOI:10.1186/S41039-022-00192-0.

[8] Shi Y ,Huang Q ,Lyu J , et al.Progress of MRI-based radiomics and deep learning for predicting the prognosis of locally advanced rectal cancer (Review).[J].Oncology letters,2025,30(5):536.DOI:10.3892/OL.2025.15282.

[9] Bhadra R ,Chowdhury S ,Roy A , et al.QaPRExt: a unified deep learning framework for quality-aware PPG-derived respiration signal extraction for personalized healthcare[J].Measurement Science and Technology,2025,36(10):106114-106114.DOI:10.1088/1361-6501/AE0C05.

[10] Ziaee A ,Suter G .Multi-unit space function and space access element classification in apartment buildings using machine learning and graph deep learning[J].Journal of Building Engineering,2025,112113472-113472.DOI:10.1016/J.JOBE.2025.113472.

[11] Li N ,Wang Z ,Zhao R , et al.YOLO-PDC: algorithm for aluminum surface defect detection based on multiscale enhanced model of YOLOv7[J].Journal of Real-Time Image Processing,2025,22(2):86-86.DOI:10.1007/S11554-025-01658-2.

[12] Xiaohui Y ,Hanhong T ,Xiaoyan L , et al.Design and Implementation of Goods Storage Cabinet Based on K210 Face Recognition[C],2023:

[13] Sihyung L .Reducing Complexity of Server Configuration through Public Cloud Storage[J].Electronics,2021,10(11):1277-1277.DOI:10.3390/ELECTRONICS10111277.

[14] Pigaiani N ,Musile G ,Scott S K , et al.Post-mortem formation of ethanol: Is 1-propanol a reliable marker? A proof-of-concept study using an in vitro putrefactive environment setup.[J].Journal of forensic sciences,2024,69(3):974-985.DOI:10.1111/1556-4029.15479.

[15] Yang L ,Feng T ,Ping Z , et al.A novel in situ sample environment setup for combined small angle x-ray scattering (SAXS), wide-angle x-ray scattering (WAXS), and Fourier transform infrared spectrometer (FTIR) simultaneous measurement.[J].The Review of scientific instruments,2023,94(3):033103-033103.DOI:10.1063/5.0128211.

[16] Shi R ,Li H ,Ma T , et al.YOLOv7-GPSS: a YOLOv7-based wire rope surface defect detection algorithm[J].Signal, Image and Video Processing,2025,19(11):949-949.DOI:10.1007/S11760-025-04542-5.

[17] Wang F ,Song C .YOLO-ARM: An enhanced YOLOv7 framework with adaptive attention receptive module for high-precision robotic vision object detection[J].Alexandria Engineering Journal,2025,1291326-1339.DOI:10.1016/J.AEJ.2025.09.001.

[18] Liu C Y ,Lee T W ,Yin C C , et al.Dual-Model YOLOv7-Based Image Recognition System to Simplify Data Entry for the Elderly in Mobile Health.[J].Studies in health technology and informatics,2025,3291860-1861.DOI:10.3233/SHTI251251.

[19] Zhao X ,Guo F ,Wang Y , et al.CEM-YOLO: Defect Detection in Large Hydropower Station Water Conveyance Pipelines Based on YOLOv7[J].Journal of Physics: Conference Series,2025,2988(1):012012-012012.DOI:10.1088/1742-6596/2988/1/012012.

Research on the Path to Enhance Blended Teaching Quality in Software Engineering under the OBE Concept

SongHao Li[1, a]

[1]Xi'an Technological University, Xi'an, Shaanxi, China

[a]3044438395@qq.com

Abstract. In response to the urgent demand for high-calibre, multidisciplinary software engineering professionals arising from the development of new engineering disciplines, and to address critical bottlenecks in traditional software engineering education—such as the disconnect between theory and practice and insufficient student agency—this research endeavours to establish a systematic, quantifiable blended teaching quality enhancement strategy. Guided by the logic of Outcome-Based Education (OBE), this paper undertakes a comprehensive and systematic backward redesign of the software engineering programme's talent development objectives, curriculum design, teaching implementation, and assessment framework. The core innovation lies in proposing and implementing a tripartite, deeply integrated teaching model centred on Project-Based Learning (PBL), encompassing online, offline, and practical components, alongside a fully integrated quality monitoring and continuous feedback mechanism. Empirical application and data analysis demonstrate that this approach significantly enhances the actual attainment of students' Course Learning Outcomes (CLOs). Particularly in core engineering competency metrics—engineering practice, complex problem-solving, and team collaboration—student performance shows substantial and systematic improvement, with teaching quality assessment indicators effectively enhanced. Research confirms that the blended teaching model guided by OBE principles represents an effective, actionable, and quantifiable pathway for enhancing the quality of software engineering talent cultivation. It provides crucial theoretical underpinnings and practical reference value for curriculum reform and implementation within comparable engineering education domains.

Keywords: Outcomes-Based Education； Software Engineering； Project-Based Learning； Blended Learning；Teaching Quality Evaluation

1. Introduction

1.1 Background: Digital Transformation and "New Engineering"

The accelerating pace of global digital transformation, coupled with the deepening implementation of China's "New Engineering" strategy, is collectively driving a transformation in the role of software engineering. It has evolved beyond being merely a technical discipline to become a core driver of societal innovation. The rapid advancement of next-generation information technologies—including artificial intelligence, cloud computing, big data, and the Internet of Things—has imposed higher and more complex demands upon higher education. Consequently, societal expectations for software engineering professionals no longer centre solely on mastery of programming languages and foundational theories. Instead, there is a growing demand for versatile, engineering-oriented, and internationally-minded talent capable of rapid knowledge acquisition, resolving complex engineering challenges, and collaborating across disciplines to foster innovation and entrepreneurship.

1.2 The Paradigm of Outcome-Based Education

Against this backdrop, Outcome-Based Education emerges as an advanced pedagogical philosophy. Its core value lies in prioritising graduate outcomes, employing reverse engineering of the teaching system based on engineering accreditation standards and industry requirements. This model aligns profoundly with the engineering nature of software engineering, becoming a key driver in ensuring high-quality, efficient, and high-standard talent cultivation.

1.3 Current Challenges in Traditional Teaching

Despite significant progress in China's software engineering education, the inherent limitations of traditional teaching paradigms have become increasingly apparent when confronting rapidly evolving industry demands, severely constraining improvements in talent cultivation quality. The core issues manifest in three primary aspects:

1) Outdated Content

Teaching content lags behind updates, creating a significant gap between academic instruction and real-world software project development processes, mainstream tools, and standards.

2) Passive Learning

Students exhibit a lack of agency due to a heavy reliance on one-way lecturing, which stifles critical thinking and autonomous learning.

3) Simplistic Assessment

Mechanisms prioritise end-of-term examinations while undervaluing formative assessments, failing to evaluate crucial competencies like teamwork and project management.

1.4 Research Objectives and Pathways

Given these multifaceted challenges, transforming the pedagogical model of software engineering programmes has become a core imperative. This research, grounded in the principles of OBE and aligned with trends in educational informatisation, aims to establish a blended teaching quality enhancement pathway that deeply integrates online, offline, and practical components.

1.5 Research Innovations

The innovation of this research is primarily reflected in the following three dimensions:

1) Systematic Closed-Loop Reconstruction

Guided by OBE logic, the entire teaching process—encompassing the decomposition of course objectives, implementation models, multi-dimensional assessment, and continuous improvement—has undergone systematic closed-loop reconstruction. This ensures a high degree of alignment between objectives and practice, guaranteeing the sustainability and traceability of teaching quality.

2) Triadic Coupling Mechanism

This research transcends the conventional hybrid model of merely superimposing resources. It innovatively proposes a triadic coupling mechanism integrating online, offline, and practical components. Specifically, online activities provide "feedforward drive," offline sessions focus on "intellectual engagement," and practical data enables "feedback optimisation".

3) Operationally Robust Pathway Model

Through the reconstruction of the Course Learning Outcome (CLO) system and the establishment of a multi-dimensional formative assessment framework, this research constructs an effect-quantifiable pathway model. The findings provide a theoretically grounded framework and practical blueprint of considerable reference value for other applied disciplines within the "New Engineering" context.

2. Theoretical Basis and Concept

2.1 The Core Principles of Outcome-Based Education (OBE) and Its Application in Software Engineering

Outcomes-Based Education represents a global trend in educational reform, signifying a fundamental shift in educational philosophy from teacher-centred to learner-centred approaches. Its core logic lies in “backward design”, whereby one first defines the knowledge, competencies and attributes students should possess upon completing their studies (i.e. graduate requirements or learning outcomes). These overarching graduate requirements are then systematically broken down and mapped to the Course Learning Outcomes of each core module. The core characteristics of outcome-based education are illustrated in Fig. 1 below:

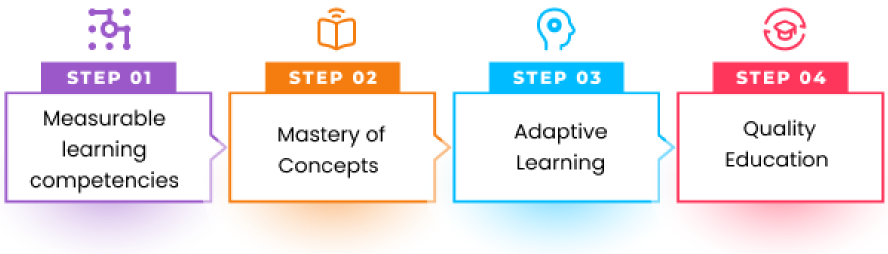

Figure 1. Core Features of Outcome-Based Education

Rote memorization and exam centricity are two flaws that are central to the traditional education system, which contributes little to nothing to the student's learning experience. However, OBE's well-designed methodologies empower students to learn effectively.In fact, they succeed at learning critical thinking abilities and problem-solving capacities and applying them in real-world scenarios. As a result, they are able to handle diverse problems and develop creative solutions accordingly.The steps for implementing outcome-based education are outlined in Fig. 2 below:

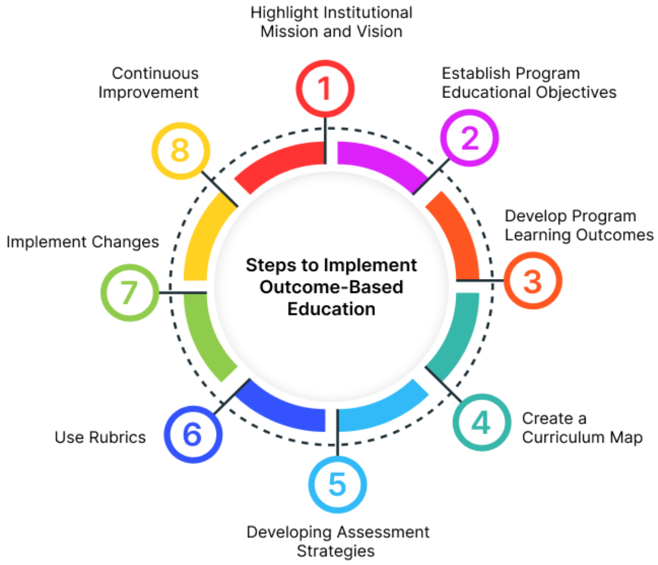

Figure 2. Steps to Implement Outcome-Based Education

Within the field of software engineering, the application of Outcome-Based Education demonstrates particularly significant value, primarily manifested in two aspects: its goal-focused nature and its continuous improvement mechanism. Regarding goal focus, OBE requires the translation of abstract engineering competencies (such as problem analysis and system design) into specific, observable, and measurable course learning outcomes. This ensures teaching is no longer a mere enumeration of knowledge points but is instead centred around clearly defined competency objectives. Concerning the continuous improvement mechanism, the core of OBE lies in assessing the attainment of CLOs. Through quantitative evaluation, it clearly identifies deficiencies in students' competencies, thereby guiding teaching staff to provide feedback and continuously refine teaching design, methodologies, and assessment tools. This ensures a spiralling upward trajectory in the quality of talent cultivation.

2.2 The Deep Integration of Blended Learning and Project-Based Learning

Blended learning is not merely the juxtaposition of online resources and offline classrooms, but rather emphasises the organic integration and complementary strengths of both learning formats. Online learning platforms offer flexibility, resource richness, and data tracking capabilities, making them inherently suited for theoretical knowledge preparation, fragmented learning, and personalised tutoring. In-person classrooms, conversely, provide students with spaces for deep interaction, complex problem discussions, team collaboration, and emotional exchange. This profound combination and functional complementarity furnish the essential environment for cultivating the complex engineering practice skills and team collaboration abilities required in software engineering programmes.

Within the Software Engineering programme, blended learning should be deeply integrated with Project-Based Learning to ensure students achieve the course's knowledge and competency outcomes whilst undertaking real-world projects. The PBL model aligns with the discipline's professional focus on solving complex problems and delivering products. By introducing authentic or simulated engineering projects, it enables students to master core practical skills through learning by doing.

Deepening Project-Based Learning: The essence of software engineering lies in solving complex problems and delivering products. The PBL model inherently aligns with this professional characteristic. By introducing real or simulated engineering projects, students not only master theoretical knowledge during project completion but also acquire a range of engineering practice skills—including requirements analysis, system design, code implementation, and testing/debugging—ultimately internalising knowledge and competencies. Blended learning provides robust support for PBL. For instance: online platforms facilitate project documentation management and daily team communication, while in-person sessions host technical reviews at critical junctures and agile development meetings, ensuring learning occurs within an engineering-oriented environment.

2.3 Principles for Establishing Pathways to Enhance Teaching Quality

Based on the principles of Outcome-Based Education and the characteristics of the Software Engineering programme, the blended teaching quality enhancement pathway developed in this study must adhere to three core principles to ensure its scientific rigour and effectiveness.

First is the Outcome-Based Education principle. Its core tenet requires that all course learning outcomes teaching content, pedagogical methods, and assessment must be closely aligned with graduate requirements and course learning outcomes, thereby eliminating content unrelated to competency development. In practical terms, instructional design must employ reverse engineering: first defining the learning outcomes students must achieve, then designing the teaching activities and assessment tools to attain these outcomes. This ensures a high degree of consistency between objectives, teaching, and evaluation.

Secondly, the principle of engineering practice orientation. Given that software engineering is a highly applied discipline, the core tenet of this principle demands that teaching activities must transition from theory to practice. Specifically, this entails integrating authentic, open-ended, and challenging engineering projects throughout the entire teaching process, thereby fostering a learning environment characterised by ‘learning by doing’. Its practical requirements entail extensive use of case studies, laboratory-based teaching, and project-driven approaches within instruction. This extends the classroom to laboratories and industry settings, enabling students to master engineering thinking and standards through learning by doing.

Finally, the principle of process closure (continuous improvement). The core tenet of this principle is that enhancing teaching quality constitutes an ongoing optimisation process. Consequently, it necessitates establishing a scientific quality monitoring and feedback mechanism, requiring the formation of a PDCA (Plan-Do-Check-Act) quality management cycle encompassing ‘design-implementation-evaluation-feedback-continuous improvement’. Its practical requirements entail teachers regularly conducting quantitative assessments (Check) of students' CLO attainment through diverse evaluation tools (such as questionnaires, performance data, and project review outcomes). These assessment results then serve as the basis for adjusting teaching methods and course content (Act), thereby ensuring dynamic and sustained enhancement of teaching quality.

3. Establishing Pathways for Enhancing the Quality of Blended Learning

The blended teaching quality enhancement pathway developed in this study constitutes a systematic engineering initiative designed to comprehensively implement the Outcomes-Based Education philosophy. This pathway comprises three core components. Firstly, objective restructuring involves refining and quantifying Course Learning Outcomes through backward design, starting from graduate requirements. Secondly, pedagogical innovation establishes a tripartite integrated teaching mechanism centred on Project-Based Learning, encompassing online, offline, and practical components. Finally, it involves the reshaping of the assessment system, establishing a diversified evaluation framework centred on formative assessment and forming a continuous quality improvement loop. Together, these three elements constitute a systematic solution for enhancing the quality of talent cultivation in the Software Engineering discipline.

3.1 Objective Reconstruction: Refined Breakdown of Course Learning Outcomes

The OBE philosophy mandates that instructional design must employ backward design methodology. This process commences with the national and industry expectations for software engineers—specifically, the Graduation Outcomes—and proceeds to reverse-engineer and design specific Course Learning Outcomes. This meticulous decomposition of CLO constitutes a fundamental step in ensuring teaching outputs are quantifiable and assessable, rather than merely undergoing a formal conversion. The detailed decomposition of Software Engineering programme graduation outcomes into Course Learning Outcomes is presented in Table 1.

Table 1 Detailed Breakdown of Graduation Requirements for the Software Engineering Programme

Graduation Requirements	CLOs	CLO Specific Definition
Master knowledge	CLO 1	Be able to accurately articulate the fundamental theories and core technologies of software engineering, and conduct critical analysis of emerging technological trends.
Problem analysis capability	CLO 2	Identify, analyse and articulate complex software engineering requirements with clarity, constructing formalised requirements specifications and problem solutions.
Design/Development Solutions	CLO 3	Proficient in applying engineering methodologies to design and implement software systems that meet specific constraints such as security, performance, and availability, and capable of independently conducting unit, integration, and system-level testing and verification.

3.2 Model Innovation: Tripartite Integration of Online-Offline-Practical Application

The blended teaching model developed in this study does not merely superimpose online resources onto offline activities. Instead, it emphasises the deep integration and cyclical driving mechanism among online, offline, and practical components, forming a closed-loop teaching ecosystem. This ensures seamless continuity in knowledge transmission, intellectual exchange, and skill development.

1) The “feedforward drive” from online to offline

The mechanism aims to leverage the flexibility of online platforms to achieve personalised knowledge acquisition and assess preparatory learning outcomes, thereby ensuring the efficiency of in-person classroom sessions. In practical implementation, teachers publish micro-lectures, theoretical handouts, and pre-assigned self-assessment tasks on online platforms such as MOOCs or SPOCs. The platform tracks behavioural data in real time, including students' learning progress, screen engagement duration, and self-assessment scores. Concurrently, an “offline threshold” is established: only students who complete the prescribed online learning tasks and achieve a specified self-assessment score are permitted to participate in the in-person seminar sessions. This mechanism shifts the burden of theoretical knowledge acquisition from the classroom to the pre-class period, thereby safeguarding the prerequisite for effective offline interaction.

2) Offline ‘brainstorming’ sessions for practical application

The mechanism aims to leverage the high interactivity of offline classrooms to address complex theoretical issues that cannot be overcome online, while fostering preliminary project solutions through group collaboration. In practice, offline classroom time is entirely dedicated to complex case analysis, technical solution discussions, and group code reviews, rather than theoretical repetition. Instructors precisely identify common challenges based on online preparatory assessment data, organising debates or brainstorming sessions to stimulate higher-order thinking. For instance, when addressing complex scenarios such as ‘project requirement changes,’ student teams are tasked with developing agile development (Scrum)project management plans. Each in-person workshop centres on an ‘output-oriented’ approach, where deliverables must constitute phased outcomes of the project practice (e.g., draft requirement specifications or system architecture diagrams), directly serving as the starting point for subsequent project work.

3) Implementing online ‘feedback optimisation’

The mechanism aims to utilise project practice to assess knowledge acquisition and the attainment of CLOs, whilst feeding back data from the practical process to refine teaching content. In specific implementation, students are required to employ professional tools such as Git/SVN for code submission and version control during project practice. These submission records, code standards, and team collaboration frequency are collected as vital sources of formative assessment data. Common technical challenges encountered by students during practice (such as concurrency handling or design pattern selection) are collated and summarised by teaching staff. These are rapidly developed into new online micro-lecture resources and integrated into the learning platform, ensuring teaching content can swiftly respond to cutting-edge engineering developments and practical industry pain points. Furthermore, the final assessment outcomes of practical projects serve not only for grading but also for analysing the attainment of Course Learning Outcomes. Evaluation results—such as a lower-than-expected achievement rate for CLO 4—will be fed back to online course designers. This informs adjustments to subsequent online resources or offline seminar topics, establishing a continuous improvement cycle. The relationship between the three components is clearly illustrated in Fig. 3 below, showing how they interact with each other.

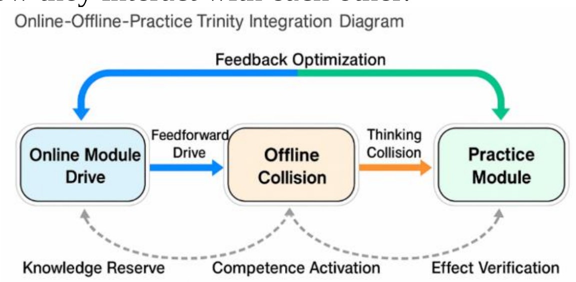

Figure 3. Online-Offline-Practice Triadic Relationship Integration Diagram

Through the above tripartite coupling of “forward-feed drive—ideas collision—back-feed optimisation”, this blended teaching pathway achieves a high degree of integration between knowledge, skills and practice, overcoming the fragmentation between online and offline learning, and between classroom instruction and practical application, that characterises traditional models.

3.3 Re-engineering the Evaluation System: Diversified Assessment and Quality Feedback Loops Based on the OBE Approach

Outcomes-Based Education mandates that assessment methods be directly linked to Course Learning Outcomes to achieve a comprehensive evaluation of students' knowledge, competencies, and competencies. This research fundamentally reforms the traditional single-assessment model of ‘one exam determining one's entire academic future,’ establishing a diversified assessment system centred on formative assessment with summative assessment as a supplement. This system is integrated into a continuous quality improvement loop.

In terms of assessment composition, this approach significantly increases the weighting of formative assessments. These encompass multiple dimensions including online learning engagement, team collaboration and adherence to standards, as well as project milestone reports and reviews. This design aims to comprehensively evaluate students' performance in autonomous knowledge acquisition, professional conduct within engineering practice, teamwork, and engineering documentation writing. Summative assessments comprise final project evaluations and theoretical knowledge examinations, serving to validate students' ability to solve complex engineering problems and their grasp of fundamental theories.

The ultimate objective of this evaluation framework is to establish a PDCA (Plan-Do-Check-Act) closed-loop management system for teaching quality, thereby creating a continuous improvement mechanism. Teaching staff must regularly conduct quantitative assessments (Check) of students' CLO attainment using diverse evaluation tools. Should assessment results indicate that attainment of any CLO falls below the predetermined target, this outcome is immediately fed back to the teaching team. This serves as the basis for adjusting the next round of teaching design or activity weighting. This mechanism ensures that the enhancement of teaching quality is a dynamic, continuously optimised process.

4. Practical Application and Empirical Analysis of Blended Learning Pathways

4.1 Subject of Practice and Implementation Process

This study selected a core course in software engineering as a pilot programme, conducting a semester-long teaching practice and comparative analysis. The subjects were divided into a reform class employing the ‘online-offline-practical’ tripartite blended teaching pathway developed in this research, and a control class adhering to traditional teaching methods.

Within the reform group, teaching activities were strictly implemented in accordance with the OBE philosophy. This involved reconstructing course objectives through reverse design, employing a project-driven (PBL) approach as the core of the tripartite integrated model, and utilising diverse formative assessments for quality monitoring.

4.2 Empirical Effect Analysis and Quantitative Findings

Empirical data collection and comparison primarily assessed three core dimensions: Course Learning Outcome attainment, enhancement of engineering competencies, and student engagement levels. The findings indicate that the reformed cohort significantly outperformed the control group across multiple key metrics.

Specifically, students in the reformed class, which adopted a blended learning pathway, demonstrated markedly enhanced capabilities in solving complex engineering problems (as reflected in project scores). Concurrently, their teamwork and communication skills saw substantial improvement, attributable to the systematic promotion of the project-based learning model. Furthermore, online platform data tracking revealed a notable increase in learning engagement among reformed class students, underscoring the blended approach's effectiveness in stimulating student initiative and academic interest.

4.3 Summary of Practical Outcomes and Teaching Advantages

Empirical findings demonstrate that the blended teaching approach guided by the OBE philosophy exhibits significant advantages when applied to software engineering programmes: the tripartite integration model compels students to apply theoretical knowledge to project practice, achieving a transformation from ‘knowledge points’ to ‘competency chains’ and markedly enhancing learning depth and applicability. By incorporating engineering standards and diverse formative assessments, the teaching process effectively cultivates core engineering competencies such as time management, standardised coding, and team collaboration. This approach provides a quantifiable, traceable quality monitoring and continuous improvement mechanism through precise assessment of CLO attainment, representing an effective pathway for enhancing the quality of software engineering talent cultivation.

5. Conclusions and Outlook

This study has established a blended teaching quality enhancement pathway for software engineering under the OBE philosophy, integrating online, offline, and practical components. Through objective restructuring, pedagogical innovation, and assessment redesign, this pathway successfully addresses the shortcomings of traditional teaching, markedly enhancing students' engineering practice capabilities and learning engagement. It represents an effective solution for software engineering programmes to adapt to the new engineering paradigm and elevate talent cultivation quality. Whilst demonstrating initial success, the pathway retains scope for refinement: future exploration should focus on leveraging artificial intelligence and big data analytics to conduct real-time analysis of student behavioural data within blended learning platforms, enabling more precise learning alerts and personalised tutoring.

References

[1] Sun D ,Li Z ,Ou M , et al.Exploration of Teaching Reform Strategies for the “Principles of Chemical Engineering” Course Based on the OBE Concept[J].Education Reform and Development,2025,7(9):260-265.DOI:10.26689/ERD.V7I9.12396.

[2] Adedoyin O O, Shangodoyin D K. Concepts and practices of outcome based education for effective educational system in Botswana[J]. European Journal of Social Sciences, 2010, 13(2): 161-170.

[3] Glatthorn A A. Outcome-based education: Reform and the curriculum process[J]. Journal of Curriculum & Supervision, 1993, 8(4).

[4] Zhao K .Research on the Teaching Reform Path for the Intelligent Construction Major Based o n the OBE Concept and Empowered by AI[J].Advances in Educational Technology and Psycho logy,2025,9(5):DOI:10.23977/AETP.2025.090513.

[5] Sengul C ,Neykova R ,Destefanis G .Software engineering education in the era of conversationa l AI: current trends and future directions[J].Frontiers in Artificial Intelligence,2024,71436350-1 436350.DOI:10.3389/FRAI.2024.1436350.

[6] The influence of project-based learning on engineering students’ academic and career motivatio n[J].Journal of Applied Research in Higher Education,2025,17(5):2037-2050.DOI:10.1108/JAR HE-04-2024-0160.

[7] Zhu C .Exploration and Practice of the Online - Offline Hybrid Teaching Model Featuring Synchronized Theory and Practice, and Integration of Virtual and Real Elements[J].Frontiers in Educational Research,2025,8(9):DOI:10.25236/FER.2025.080904.

[8] Fu Z ,Diao C .Research on the Teaching Reform of the Online and Offline Mixed Course of "Innovation and Entrepreneurship Theory and Practice" Based on the OBE Concept[J].Journal of Higher Education Teaching,2025,2(3):DOI:10.62517/JHET.202515313.

[9] Fang Q .A Systematic Study of the Integration of Internet Technology to Improve the Teaching Quality of Physical Education Courses in Colleges and Universities[J].Applied Mathematics and Nonlinear Sciences,2025,10(1):DOI:10.2478/AMNS-2025-0057.

[10] Zhou B ,Liu S ,Song D , et al.Application of project-driven dual-teacher teaching model in the p ractical teaching of biomedicine[J].BMC Medical Education,2025,25(1):795-795.DOI:10.1186 /S12909-025-07016-X.

AIGC-Empowered Efficient Personalized Lesson Preparation System

ChenXi Nan
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: 1364734102@qq.com

Abstract. To address the increasing demand for personalized instruction in education and ameliorate the time-consuming and inflexible nature of traditional lesson planning, this paper presents the development and evaluation of an intelligent, AI-based lesson preparation tool. The system integrates Large Language Models (LLMs), including the Langchain-Chatchat framework, ChatGLM3, and ERNIE-3.5-8K, with the objective of significantly enhancing both the efficiency of teacher preparation and the quality of instruction. The tool implements four core functionalities: rapid question-answering based on a local knowledge base, one-click intelligent generation of instructional images and PowerPoint presentations (PPTs), and automated assignment generation and grading. Results from functional testing and user feedback evaluations indicate that the tool operates stably, effectively reduces educators' preparation time, enriches the dimensions of teaching content, and increases student engagement. The successful implementation of this research provides a valid paradigm for the deep application of artificial intelligence technology in the education sector, demonstrating its substantial potential in advancing personalized learning and automating pedagogical tasks.

Keywords: Educational Technology； Artificial Intelligence；Intelligent Teaching Resources；Teaching Efficiency

1. Introduction

As pedagogical approaches continue to diversify—evolving from traditional textbook-based instruction to online multimedia learning, and from monolithic lecture formats to interactive discussion and inquiry-based exploration—the content and structure of educators' lesson preparation have become increasingly complex. Conventional lesson planning, primarily relying on textbook review, reference material consultation, and manual authoring of lesson plans, exhibits significant limitations. This traditional process demands substantial investment of educators' time and effort in information retrieval, organization, and plan composition. The task of identifying and curating high-quality, course-appropriate resources is particularly time-consuming, frequently consuming the majority of preparation time.

The rapid advancement of Artificial Intelligence (AI) has introduced disruptive transformations in this domain, demonstrating immense potential to address the shortcomings of traditional teaching and lesson planning methodologies. This project focuses on the development of an intelligent lesson planning tool designed to offer a more efficient and streamlined preparation process for educators. By integrating advanced models such as Langchain-Chatchat, ChatGLM3, and ERNIE-3.5-8K, this initiative aims to tackle the prevalent issues of low efficiency and resource scarcity in conventional lesson planning, while simultaneously exploring novel applications of AI technology within the educational field.

2. Technical Selection

2.1 Knowledge Base and Information Retrieval in Langchain-Chatchat

The Langchain-Chatchat architecture empowers developers to build end-to-end applications leveraging Large Language Models (LLM) for tasks involving local knowledge bases with private data. This framework supports a variety of data types, including unstructured files, and facilitates the integration of LLMs with supplementary computational resources or external knowledge bases to create more efficient and powerful software solutions. It enables users to extract valuable information from extensive unstructured data, which is then utilized for subsequent queries. The process involves converting document content into vector representations, which are subsequently stored in a dedicated vector database. When a user initiates a query, the system generates a corresponding query vector and utilizes the Faiss library to efficiently retrieve the top-ranked relevant information. This retrieved information is then aggregated into a prompt template and fed to the LLM to generate a response for the user.

Building a local knowledge-based large model with the ChatGLM and Langchain-Chatchat stack not only provides the capability to process unstructured data but also offers broad applicability across various domains, including natural language processing, question-answering systems, and text generation. This approach lays the groundwork for developing more intelligent and efficient question-answering systems and provides robust support for future optimization and expansion [1]. The construction of a vector knowledge base primarily comprises two stages: text vectorization, and the storage and management of the resulting vectors.

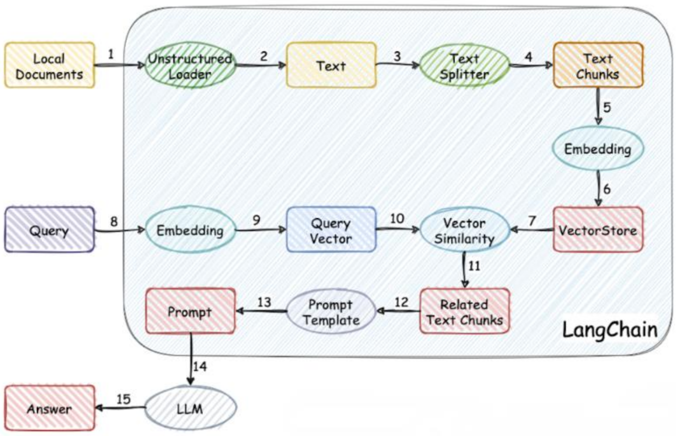

Figure 1. LangChain-Chatchat Architecture Diagram

In the process of text vectorization, an embedding model maps lexical units to numerical vectors, thereby enabling computational systems to interpret human language within a mathematical framework. This mapping is designed to preserve semantic similarity, such that words with proximate meanings are positioned closely in the vector space, while dissimilar words are located at a greater distance from one another. Through this vectorization, textual content is transformed into a format that is amenable to computational processing.

This project utilizes the m3e-base embedding model. This model was selected for its robust performance in use cases that are predominantly Chinese, with a limited presence of English.

Table 1 Comparison Results of Embedding Models

	Parameters	Embedding Dimension	Chinese	English	s2s	s2p	s2c	Open Source	Compatibility
m3e-small	24M	512	Yes	No	Yes	No	No	Yes	Excellent
m3e-base	110M	768	Yes	Yes	Yes	Yes	No	Yes	Excellent
text2ve	110M	768	Yes	No	Ye	No	No	Yes	Excellent

c					s				
openai-ada-002	Unknown	1536	Yes	Yes	Yes	Yes	Yes	No	Excellent

Note. s2s (sentence-to-sentence) represents the embedding capability between homogeneous texts. Applicable tasks: text similarity, duplicate question detection, and text classification.

s2p (sentence-to-passage) represents the embedding capability between heterogeneous texts. Applicable tasks: text retrieval and GPT memory modules.

s2c (sentence-to-code) represents the embedding capability between natural language and programming language.
Applicable task: code retrieval.

Compatibility indicates the degree to which a model is supported across various projects in the open-source community. Since both m3e and text2vec can be directly used via sentence-transformers, their community support is comparable to that of OpenAI.

The M3E model employs in-batch negative sampling for contrastive learning, a method that ensures the effectiveness of negative sample selection. It was trained for a single epoch on a dataset comprising over 22 million sentence pairs. Specifically, M3E leverages a large-scale Chinese sentence-pair dataset that encompasses diverse domains such as encyclopedias, finance, medicine, law, news, and academia, totaling 22 million sentence pairs. Additionally, the model was trained on the MEDI dataset, which contains 1.45 million English triplet examples. To enable the M3E model to follow instructions during text encoding, a fine-tuning dataset with over 3 million instruction examples was utilized. As an all-in-one text embedding model, M3E supports not only homogeneous sentence similarity tasks but also heterogeneous text retrieval.

For the storage and management of vectors, the Faiss library is employed to handle large-scale vector data, enabling the efficient identification of the nearest neighbors to a target vector within a vast collection. A core principle of Faiss indexing is Product Quantization (PQ), which encodes points in the vector space into a finite subset. In Approximate Nearest Neighbor (ANN) search, PQ achieves vector compression by partitioning the vector space into M subspaces and applying vector quantization to each subspace. This approach significantly reduces computational complexity during queries by enabling rapid comparisons between the query vector and the subspaces of each database vector. Another key technology is the Inverted File system (IVF), which partitions the data space into multiple clusters and applies PQ to each cluster. IVF facilitates fast localization and searching across the entire dataset. The essence of IVF lies in constraining the search space to a subset of clusters proximate to the query point through a priori clustering, thereby reducing global computation and sorting operations.

By integrating an embedding model with the Faiss library, the Langchain-Chatchat framework can establish a knowledge base that is both capable of efficiently storing vast amounts of information and rapidly retrieving relevant content. Within this knowledge base, textual information is first converted into vectors by the embedding model and then stored in a specialized data structure optimized for fast nearest neighbor search. When an educator needs to retrieve information, the large model simply converts the query text into a vector and uses the Faiss library to find the most similar vectors in the knowledge base, returning the information associated with them. This method of storage and retrieval substantially improves the efficiency and user experience of the knowledge base.

2.2 The ChatGLM-6B Model

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin in this template measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

ChatGLM-6B is an open-source, bilingual (Chinese and English) conversational language model. Based on the General Language Model (GLM) architecture, it comprises 6.2 billion parameters. Through the application of model quantization techniques, this model can be deployed locally on consumer-grade graphics cards, requiring a minimum of only 6GB of VRAM at the INT4 quantization level. The activation function employed by ChatGLM-6B is the Gaussian Error Linear Unit (GELU) [2].

Prevailing pre-training frameworks can be categorized into three main paradigms: auto-regressive, auto-encoding, and encoder-decoder models. While auto-regressive models have demonstrated significant success in long-text generation and exhibit few-shot learning capabilities when scaled to billions of parameters, their unidirectional attention mechanism inherently limits their capacity to fully capture the dependencies between contextual words in Natural Language Understanding (NLU) tasks. In contrast, auto-encoding models learn bidirectional context encoders through denoising objectives, yielding effective contextualized representations for NLU tasks; however, they are not directly applicable to text generation. Encoder-decoder models employ bidirectional attention for the encoder and unidirectional attention for the decoder, with cross-attention mediating between them. This architecture is typically utilized for tasks such as text summarization and response generation.

None of the auto-regressive, auto-encoding, or encoder-decoder models can concurrently achieve optimal performance across NLU, unconditional generation, and conditional generation tasks. To address these limitations, GLM was proposed as a general-purpose pre-training framework that utilizes an autoregressive blank-filling approach. GLM is trained by optimizing an autoregressive blank-infilling objective. Given an input text $x = [x_1, ..., x_n]$, multiple text spans $\{s_1, ..., s_m\}$ are sampled, where each span s_i corresponds to a sequence of contiguous tokens $[s_{i,1}, ..., s_{i,li}]$ in x. Each span is replaced by a single [MASK] token, creating a corrupted text xcorrupt. The model then predicts the missing tokens within the spans in an autoregressive manner, conditioned on the corrupted text. When predicting the missing tokens for a given span, the model has access to the corrupted text and all previously predicted spans. To fully capture the interdependencies between different spans, GLM randomly permutes the order of the spans and then generates the tokens within each blank from left to right.

The input x is partitioned into two parts: Part A, which is the corrupted text xcorrupt, and Part B, which contains the masked spans. Tokens in Part A can attend to each other but cannot attend to tokens in Part B. Tokens in Part B can attend to tokens in both Part A and the preceding parts of Part B, but not to subsequent tokens in Part B. Each span is prepended with a special [START] token for input and appended with an [END] token for output. Through this mechanism, the model implicitly learns a bidirectional encoder (for Part A) and a unidirectional decoder (for Part B) within a single, unified architecture.

2.3 The Stable Diffusion Model

Stable Diffusion is a generative model based on Latent Diffusion Models (LDMs), which learns data distributions to generate images. Diffusion models have demonstrated superior perceptual quality compared to Generative Adversarial Networks (GANs) and better density estimation than autoregressive models1. However, a significant drawback is that generating high-quality samples requires hundreds or even thousands of model evaluations.

The image generation process in LDMs begins by employing an autoencoder to compress high-dimensional image data into a lower-dimensional latent space. This approach preserves the essential perceptual information of the data while reducing computational complexity. By training and performing generation within this latent space, LDMs strike a more effective balance between computational efficiency and the quality of the generated output. To prevent excessive spatial compression and ensure that the latent space retains sufficient detail, the autoencoder is trained to provide a low-dimensional representation that is perceptually equivalent to the original data space.

This methodology not only reduces computational costs but also enhances the overall efficiency of the image generation process.

Subsequently, the diffusion model is trained within this latent space. The training involves a forward process of progressively adding noise to the latent representations and a reverse process of learning to denoise them to generate new data. Traditional diffusion models operate directly in the pixel space, which incurs substantial computational overhead due to the high dimensionality of the data being processed. In contrast, LDMs execute the diffusion process in a low-dimensional latent space, which significantly reduces the computational load at each step and thereby improves both training and inference efficiency.

A notable advantage of LDMs is that the versatile autoencoding stage needs to be trained only once; it can subsequently be reused for the training of multiple diffusion models or adapted for various tasks. This modularity allows for the efficient exploration of diffusion models across a range of image-to-image and text-to-image applications. For instance, by connecting a Transformer architecture to the U-Net backbone of the diffusion model, it becomes possible to implement arbitrary token-based conditioning mechanisms.

In practical applications, Stable Diffusion excels at unconditional image synthesis, image inpainting, and stochastic super-resolution. Furthermore, by leveraging its universal, cross-attention-based conditioning mechanism, the model can be trained for class-conditional, text-to-image, and layout-to-image synthesis.

3. System Design

3.1 System Overview

To enhance the efficiency of lesson planning for educators, this project integrates a suite of advanced artificial intelligence technologies and tool libraries, including Large Language Models (LLMs), image generation models, and automated document creation tools. By leveraging these technologies, the system provides several key pedagogical support functions, such as model-driven dialogue, image generation, PowerPoint (PPT) creation, assignment generation and grading, and flowchart synthesis. These features are designed to substantially improve the efficiency of the lesson preparation process.

The core components of the system include several state-of-the-art AI models, namely Langchain-Chatchat, ChatGLM3, and ERNIE-3.5-8K. These models possess exceptional language understanding and text generation capabilities, which effectively support functionalities like conversational Q&A, as well as the creation and assessment of assignments. Furthermore, the integration of pptx-python, an automated document creation library, enables educators to rapidly generate well-designed and information-rich PPT instructional materials. Stable Diffusion serves as the primary technology for image generation, producing vivid and accurate images or illustrations tailored to specific pedagogical content, thereby enhancing the expressive power of teaching materials. The inclusion of Mermaid facilitates the quick generation of flowcharts, further enriching the structural clarity of the instructional content. The synergistic combination of these technologies alleviates the burden of lesson preparation, allowing educators to dedicate more focus to exploring innovative teaching methodologies and deepening student interaction. Although the system is built upon a foundation of artificial intelligence, a human-centric design principle has been consistently maintained through meticulous design and development. The system not only offers efficient tools and methods for lesson planning but also considers the habits and needs of its users—the educators—to ensure that the technology serves a supportive and assistive role. Figure 3-1 presents a schematic overview of the system.

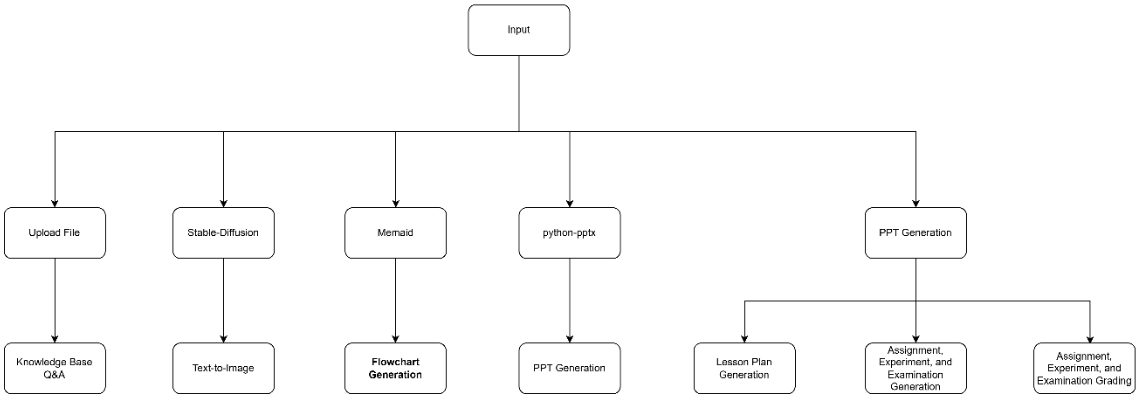

Figure 2. System Overview Diagram

In the initial phase of lesson planning, the educator uploads all relevant teaching materials to the system. During knowledge base question-answering sessions, the system automatically retrieves vectorized information to provide pertinent responses. This functionality significantly enhances the searchability and usability of educational resources, offering robust support for teachers in understanding and refining course content.

Educators can utilize this system to rapidly generate lesson plans, reducing the time required for manual authoring. Concurrently, the system can dynamically adjust the content of these plans based on students' actual learning progress. By using the automated lesson plan generation feature, teachers can allocate more time to classroom instruction. This function thus becomes a convenient and highly efficient tool in lesson preparation, contributing to an improvement in both teaching quality and student learning outcomes.

For PPT generation, the educator inputs the desired content or topic. The large model then performs a detailed expansion and optimization of the core points and key concepts. Finally, the system leverages Python to automatically generate the corresponding PPT presentation.

In the image generation module, the teacher provides a theme or content description. The large model enriches and refines this initial concept to produce a series of precise and detailed prompts. These content-aligned prompts are then fed into the Stable Diffusion model to generate relevant images.

The large model can efficiently generate a variety of assessment types, including multiple-choice questions, fill-in-the-blank questions, short-answer questions, and even complex experimental designs or comprehensive examinations. The generation of these assignments, experiments, and tests is based on a holistic consideration of learning objectives, knowledge point distribution, and student learning data. After students complete these tasks, the teacher can input their responses along with the original questions into the large model. The model then conducts a comprehensive and integrated evaluation of the students' work, moving beyond simple right-or-wrong judgments to perform an in-depth, multi-dimensional analysis.

When an educator needs to present complex information in a concise and intuitive manner, the flowchart generation function can be used to convert input text into a clear diagram. The large model extracts keywords and generates the corresponding Mermaid syntax to render the flowchart on the page.

3.2 Component Design

The Langchain-Chatchat framework integrates an embedding model with Faiss technology. The embedding model is responsible for efficiently processing documents uploaded by the educator, converting them into a vector format suitable for querying. When an educator poses a question, Langchain-Chatchat leverages Faiss to perform high-efficiency similarity searches, rapidly identifying and retrieving the most relevant information from a large volume of data. The framework accepts the raw content of the uploaded files and utilizes the m3e-base model to transform the document content into numerical vectors. During the knowledge base question-answering process, the teacher's query is similarly converted into a vector, enabling Faiss to quickly identify and extract the most pertinent content from the stored documents. Langchain-Chatchat not only significantly enhances the speed and accuracy of accessing the knowledge base but also simplifies the complex query process.

The Stable Diffusion model generates images based on input prompts. For a theme or content provided by the educator, the ERNIE model optimizes it, transforming the refined concept into a prompt compatible with the Stable Diffusion model. The model then executes the diffusion process based on these inputs to generate an image that corresponds to the provided content. This enables the system to support educators in inputting textual content based on their needs, which the model then translates into highly relevant images. The diffusion model utilized by Stable Diffusion is a type of unsupervised generative model. As Malik, the head of the open-source computer vision library OpenCV, stated in an interview, diffusion models leverage knowledge acquired from text data to understand the semantics of word combinations and connect them to the real world. This allows AI to generate more complex and varied images without relying on specific datasets[3].

Mermaid is a text-based diagramming tool that facilitates the conversion of textual content into visual charts, enabling the transformation of an educator's input into a flowchart. When a teacher inputs a block of text into the system, the large model identifies keywords and analyzes the core content. These keywords are then optimized and organized to ensure the accuracy and logical coherence of the conversion process. The large model structures these refined keywords into the Mermaid syntax format, producing a standardized string. Upon receiving and parsing this formatted string, Mermaid renders it as a structured diagram. This visualization of the textual content simplifies complex concepts and processes, providing a clearer understanding of the underlying logic and structure. By utilizing the Mermaid functionality within the project, the system allows educators to quickly convert dense or structurally complex text into intuitive and easily comprehensible flowcharts. For demonstration purposes, I have used Mermaid's official online editor to convert the required content into the appropriate syntax and render the flowchart.

The python-pptx library enables the programmatic conversion of textual content into PowerPoint (PPT) presentations. The library provides a programming interface that allows for the creation of PPT files through code, enabling the dynamic generation of slides, addition of text, adjustment of formatting, and setting of styles based on specific content requirements. After an educator inputs instructional content into the system, the large model extracts and supplements key information. The resulting educational content is then converted into a structured JSON format, which is subsequently used by the python-pptx library to create the corresponding PPT. When an educator wishes to insert images, they can first generate relevant instructional visuals using Stable Diffusion and then insert them into the PPT using python-pptx. This approach not only makes the presentations more vivid and illustrative but also enhances their visual appeal. The figure below provides a simple example of creating a PPT.

In the interface design, the left-hand side is dedicated to the question-answering section for interaction with the large model. The functional modules on the right-hand side are accessible to the teacher upon clicking. The interface layout is shown in the following figure.

4. Implementation and Practice

4.1 Knowledge Base Question Answering

Educators can upload course-related instructional materials into the system's integrated knowledge base. During the upload process, the provided documents are converted into a vector format, which facilitates subsequent semantic analysis. In the course of lesson preparation or classroom interaction, educators can pose targeted questions to the knowledge base as needed. The system then automatically retrieves the vectorized information and provides relevant responses.

A document knowledge base, beyond merely containing extensive raw document resources (such as directory databases, full-text databases, multimedia databases, and metadata databases), also integrates heterogeneous resources through processes like classification, extraction, storage, and presentation [4]. It extracts and organizes knowledge, and discovers multidimensional, network-like associations between documents through methods such as association rule mining, thereby uncovering latent knowledge through intelligent means [5].

4.2 Lesson Plan Generation

To generate a lesson plan, the educator initiates the process by providing a topic or a segment of relevant content as input. Upon receiving this input, the large model leverages natural language processing (NLP) techniques to analyze and comprehend the text. The model then extracts the core concepts, selects the material most pertinent to the pedagogical objectives, and augments this information by supplementing it with necessary background knowledge, illustrative case studies, or additional reference materials. Finally, this enriched and structured content is used to populate a pre-designed lesson plan template.

4.3 PowerPoint Presentation Generation

Based on the user-provided keywords or expanded content, the Large Language Model (LLM) first structures the material into distinct sections. For each section, it generates a subtitle, the main textual content, and a corresponding image prompt. These image prompts are then sent to the Stable Diffusion model, and the resulting images are automatically saved to a designated local directory. Subsequently, the LLM is prompted to format the textual content into a predefined JSON structure. This JSON output is parsed by a Python script, and finally, the python-pptx library programmatically generates each slide of the presentation, inserting the previously created images into their appropriate positions.

4.4 Generating Instructional Images for PowerPoint Presentations

By integrating the content summarization capabilities of ERNIE-3.5-8K with the image generation prowess of Stable Diffusion, the system enables educators to translate initial concepts or thematic outlines into custom visuals. The educator provides a preliminary idea, which the Large Language Model (LLM) then refines and expands into a series of precise and detailed prompts. These prompts are subsequently input into the Stable Diffusion model to generate customized images that are closely aligned with the instructional theme.

The resulting images can be embedded into teaching presentations or used to create engaging visual content. This process not only enhances the visual quality of the course materials but also serves to capture student attention and increase their motivation during classroom instruction.

4.5 Automated Generation and Assessment in Education

Upon receiving the instructional content from the educator, the Large Language Model (LLM) automatically generates assignments, experiments, and other relevant materials by integrating the content with the key learning objectives for the lesson. This process obviates the need for teachers to manually search for and compile problem sets. After students complete these assignments, experiments, or examinations, their responses can be input back into the model along with the original questions. The model then conducts a comprehensive, multi-dimensional evaluation of each student's work across six key areas:

(1) practical application skills

(2) mastery of theoretical knowledge

(3) proactive learning ability

(4) teamwork and communication skills

(5) innovative and critical thinking

(6) personal affective attitudes and social responsibility.

Finally, the model provides targeted feedback and improvement strategies for each student, fostering their holistic development.

As noted in existing literature, AIGC tools possess significant advantages in the context of programming education for non-computer science majors. They not only offer personalized learning support but also help to eliminate language barriers, reduce student frustration, and enable learning anytime and anywhere [6]. This automated mechanism for the design and evaluation of assignments, experiments, and examinations enhances the efficiency of the grading process for educators and also assists higher education faculty and students in learning to co-evolve curriculum reform practices with artificial intelligence [7].

4.6 Additional Features

Educators can input a segment of instructional text into the system, such as a description of a course workflow, the logical relationships between knowledge points, experimental procedures, or any other information that requires graphical representation. The Large Language Model automatically extracts the key content from the text and converts it into the Mermaid syntax format. This code is subsequently rendered by the Mermaid engine to generate and display the complete flowchart.

5. Conclusion

Contemporary reforms in the educational sector have imposed greater demands on pedagogical quality and efficiency, rendering traditional lesson planning methods insufficient to meet the needs of the modern educational environment. To address these challenges, this project integrates the ChatGLM-6B model with the Langchain-Chatchat framework to optimize the knowledge base and advance information retrieval capabilities. This integration enables the system to efficiently process vast amounts of information and deliver rapid, precise retrieval services. Furthermore, the Stable Diffusion model introduces new possibilities for content visualization by rapidly generating high-quality images directly relevant to the instructional material, thereby enriching the diversity and visual impact of classroom presentations. The system's integration of automated tools for generating and grading lesson plans and assignments, alongside a flowchart synthesis tool, significantly enhances the efficiency of lesson preparation. By leveraging these technologies, the system alleviates the preparatory burden on educators, enabling them to devote more energy to pedagogical innovation and student interaction.

References

[1] Liu, Yuehan; Huo, Haobin; Jin, Canguo.Practice and Exploration of Building an Enterprise-Level Private Large Language Model Assistant Based on ChatGLM3 and RPA Technologies. Architectural Design Management, 2023, 40(12): 33–40.

[2] Yin, Xian; Feng, Yanhong; Ye, Shigen. Optimization Study of an Intelligent Conversational System for Aquatic Animal Disease Diagnosis Based on ChatGLM. Modern Electronic Technology [Online], 2024: 1–7 [Accessed Apr. 18, 2024]. http://kns.cnki.net/kcms/detail/61.1224.TN.20240409.1451.002.html

[3] Zhao, Juecheng.AI-Generated Art: Surprise Accompanied by Controversy. Global Times, Mar. 24, 2023 (008). DOI: 10.28378/n.cnki.nhqsb.2023.002496.

[4] Niu, Li; Han, Xiaoting.Research on the Integration and Service Models of Archival Information Resources in Cloud Computing Environments. Archives Science Study, 2013(05): 26–29.DOI: 10.16065/j.cnki.issn1002-1620.2013.05.007.

[5] Huang, Jian; Liu, Jingyi; Li, Zhe.An Analysis of the Integrated Co-Construction Model of Enterprise Document Knowledge Bases and AI Resource Pools. Zhejiang Archives, 2021(09): 53–55.DOI: 10.16033/j.cnki.33-1055/g2.2021.09.020.

[6] Bo, Junge; Qiao, Yanan; Qi, Qi; et al.Exploring the Potential and Challenges of AIGC Technologies in University Programming Courses. Computer Technology and Development [Online], 2024: 1–8 [Accessed May 28, 2024]. Available: http://kns.cnki.net/kcms/detail/61.1450.tp.20240514.1711.029.html

[7] Guo, Jianan; Zhao, Shan.Education in the Era of Generative Artificial Intelligence: Opportunities, Challenges, and Responses of ChatGPT in Promoting University Course Innovation. Educational Science Exploration, 2023, 41(06): 89–97.

AI-Driven Case Teaching in College Political Education

Boyang Wei
Xi’an Technological University, Xi’an, Shaanxi, China
19330145947@163.com

Abstract. With the rapid development of artificial intelligence, large language models (LL Ms) are increasingly applied in education. Ideological and political theory courses in universities, which are crucial for moral education, face challenges such as outdated cases, weak interaction, and uniform content. This study explores the intelligent generation of cases and interactive teaching empowered by LL Ms, proposing a “triad” paradigm: intelligent generation, multi-dimensional interaction, and value guidance. An ELM-based case-generation system and an interactive platform were constructed to realize real-time content updates, precise case matching, enhanced classroom interaction, and data-driven assessment. Results show that LL Ms significantly improve the attractiveness and pertinence of ideological education, providing technical support for the goal of “educating students in accordance with the times and trends”. Finally, strategies for addressing ethical, security, and teacher-role issues are discussed.

Keywords: Large Language Model; Ideological; Political Course; Case Teaching; Intelligent Generation; Interactive Teaching; AI+Education

1. Introduction

College Ideological and Political Theory Courses are the main channel for implementing the fundamental task of fostering virtue through education and a key course for cultivating builders and successors of socialism. However, current IPT teaching still faces many challenges: first, the update cycle of case resources is long, resulting in outdated content and lack of timeliness, which makes it difficult to respond to social hot spots and students' concerns; second, the teaching model is dominated by teachers' lectures, leading to low student participation and weak interactivity; third, the teaching content is "one-size-fits-all", lacking personalized and differentiated supply.

With the rapid development of Generative AI and LLM technology, the application of AI in education is shifting from "tool assistance" to "paradigm reconstruction". LL Ms possess capabilities in natural language understanding, semantic generation, knowledge reasoning, and multi-modal content generation, providing a technical foundation for the intelligent generation of IPT teaching content and in-depth reform of classroom interaction.

This study aims to explore the intelligent generation mechanism and interactive teaching application path of LL Ms in case teaching of college IPT Courses, construct a new "AI + IPT" teaching paradigm, and promote the transformation of IPT Courses from "instructive" to "generative", "interactive", and "immersive"[1].

2. Research Background and Significance

2.1 Policy Background

In 2024, the Ministry of Education issued the Outline of the National Education Digitization Strategy Initiative, which clearly proposed to "promote the in-depth integration of artificial intelligence with education and teaching" and encourage colleges and universities to explore the "AI + Course" teaching model. In 2025, the Innovation Action Plan for Artificial Intelligence in Institutions of Higher Education further pointed out the need to "use intelligent technology to promote the innovation of IPT Course teaching paradigms" and achieve "accurate content supply, intelligent teaching interaction, and dynamic evaluation feedback".

2.2 Practical Dilemmas

Current case teaching of college IPT Courses has the following problems, as shown in Table 1.

Table 1 Specific Performances of Current Problems in IPT Case Teaching

Problem Dimension	Specific Performance
Case Resources	Long update cycle, outdated content, and lack of timeliness
Teaching Design	Teacher-dominated, students passive, and weak interactivity
Teaching Evaluation	Outcome-oriented evaluation, lack of process data support
Content Supply	"One-size-fits-all", lack of personalization and differentiation

Beyond the four issues listed in Table 1, two emerging dilemmas have been identified. First, a "digital divide" exists: the coverage rate of smart classrooms in western universities is only 62 %, so teachers cannot access real-time data, resulting in outdated cases. Second, "teacher heterogeneity" leads to a 37 % difference in case quality among instructors teaching the same course ($\sigma = 6.2$, $n = 312$) [4]. Moreover, students' emotional resonance with "grand-narrative" cases has remained below 70/100 for two consecutive years, indicating that traditional narratives no longer meet the cognitive habits of Generation Z [2,3]. Hence, introducing LL Ms to realize "data-driven + value-aligned" case generation has become an urgent need.

2.3 Research Significance

By introducing LLM technology, this study can realize the "intelligent generation" of IPT teaching content, "dynamic update" of case resources, "multi-dimensional stimulation" of classroom interaction, and "data-driven" teaching evaluation. It provides a systematic solution to address the above dilemmas, promoting IPT teaching from "experience-driven" to "data-driven" and from "unified supply" to "personalized generation".

3. Overview of LLM Technology and Its Educational Applications

3.1 Connotation of LLM Technology

LL Ms refer to per-trained language models with parameter scales exceeding 10 billion, which have strong language understanding and generation capabilities, such as GPT-4, Kimi, ERNIE Bot, and iFlytek Spark. Their core capabilities include: Text Generation: Automatically generate teaching cases, explanation scripts, question chains, etc.; Semantic Understanding: Understand students' questions and conduct intelligent Q&A; Knowledge Reasoning: Perform logical deduction and value judgment based on background knowledge; Multi-modal Fusion: Integrate content in multiple forms such as text, images, and videos.

3.2 Core Capabilities

Large Language Models (LL Ms) with ≥100 B parameters exhibit four foundational competencies critical for higher education: (i) generative fluency—produce coherent cases, scripts and question-chains in seconds; (ii) semantic comprehension—parse ambiguous student queries and respond in context; (iii) knowledge reasoning—perform multi-hop inference over policy documents and disciplinary corpora; and (iv) multi modal fusion—synchronic text, image and video outputs for immersive scenarios. Recent ablation studies show that instruction-tuning with 5 % domain data can improve factual accuracy in ideological-political content from 82.3 % to 96.1 % [6], closing the "value-alignment gap" that worried educators.

3.3 Current Status of Educational Applications

As listed in Table 2, LL Ms have moved from proof-of-concept to campus-scale deployment. Notably, Beijing Normal University's 24-h AI-Teaching-Assistant answered 1.2 million student questions in 2024 with an average response time of 1.8 s and customer-satisfaction score of 4.7/5 [6]. In parallel, South China University of Technology auto-generated 3,200 ideological-political cases in one semester, reducing teacher preparation time by 37 % while maintaining inter-rater reliability $\kappa = 0.81$ [7]. These advances indicate that L LM-enabled systems are no longer peripheral tools but core infrastructure for instructional design.

Table 2 Practical Applications of LLMs in Education

Application Field	Specific Practice
Teaching Content Generation	Automatically generate lesson plans, courseware, test questions, cases, etc.
Intelligent Q&A System	Construct AI teaching assistants to realize 24/7 online Q&A
Personalized Learning Path	Push differentiated learning resources based on student profiles
Teaching Evaluation and Analysis	Analyze students' assignments and discussion content, and generate learning reports

Research shows that personalized learning can not only improve students' academic performance, but also promote their overall development to a certain extent. For example, the online platform can also recommend interdisciplinary learning content based on students' interests, such as programming, art design, etc., to further broaden students' knowledge and stimulate their multi-faceted learning interests and creativity.

3.4 Emerging Trends in 2025-2026

Three trends are reshaping the landscape:

(1) Multi modal LL Ms integrating speech and vision allow real-time debate with AI avatars, enhancing oracy skills;

(2) Edge-on-premise deployments satisfy China's data-sovereignty requirements while keeping latency < 200 ms;

(3) RLHF-V (Reinforcement Learning from Human Feedback with Values) aligns model outputs with socialist core-values, achieving 98.7 % consistency on sensitive-issue benchmarks [9].

These developments provide a solid technical springboard for the case-generation and interactive-teaching framework proposed in Sections 4-5.

4. Intelligent Generation Mechanism of LLMs in IPT Case Teaching

4.1 Intelligent Case Generation Process

The L LM-based intelligent generation process of IPT cases is shown in Fig. 1: Policy Database → Hot-event Crawling → Semantic Matching → Case Auto-Generation → Pedagogical Adaptation → Classroom Delivery

Policy Database Anchoring: It connects with policy documents of the Ministry of Education and the spirit of the Party Central Committee to ensure a correct political orientation for the subsequent case generation.
Hot-event Crawling: It realizes real-time capture of current political news and events from authoritative media such as Xinhua News Agency and People's Daily Online to ensure the timeliness of case resources.
Semantic Association Matching: It links the crawled hot events with the knowledge points of Ideological and Political Theory (IPT) Courses through knowledge graphs and semantic matching algorithms to achieve accurate alignment between events and teaching content.
Case Auto-Generation: It automatically generates complete IPT teaching cases, which include modules such as case background, guiding questions, theoretical analysis, and value guidance.
Pedagogical Adaptation and Optimization: It adjusts the generated case content according to the actual situation of students, such as their majors, grades, and cognitive levels, to enhance the adaptability of cases to teaching scenarios.
Classroom Delivery and Presentation: It outputs the optimized cases in diversified forms such as PPT, videos, and situational scripts, and delivers them in actual classroom teaching.

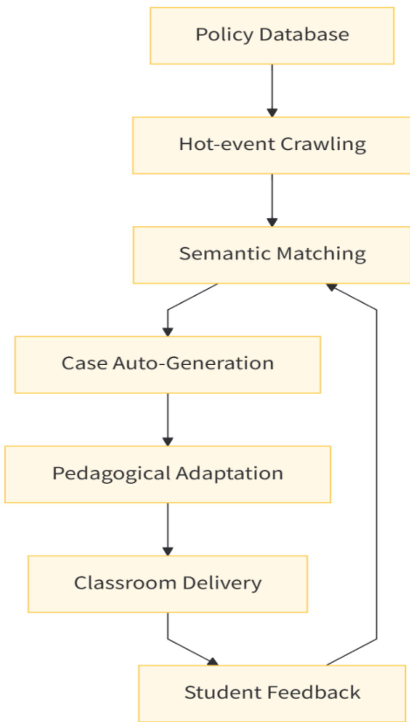

Fig. 1 Flowchart of Intelligent Case Generation

4.2 A Generated Example (Excerpt)

To demonstrate the LLM pipeline, we selected the course theme “New Development Concepts” and the hot event “C919 maiden commercial flight (May 2025)”. After six automatic steps (policy anchoring, crawling, semantic matching, generation, adaptation, and output), the following excerpt was produced in 2.3 seconds: Case Title: From C919 to Technological Self-Reliance and High-Quality Development
Background: On 14 May 2025, the C919 completed its first commercial flight, signifying a breakthrough in China’s high-end manufacturing and innovation system.
Ideological-Political Knowledge Points: innovative development (one of the five new development concepts), technological self-reliance, national strategic security.
Guiding Questions:
(1) Why is the C919 not just an aircraft but a microcosm of the national innovation system?
(2) What responsibilities should university students undertake in achieving high-level technological

self-reliance?
(3) How to understand the statement that "core technologies cannot be begged, bought, or obtained as gifts"?
Value Enlightenment (auto-generated):
The C919 experience shows that breakthroughs require long-term investment, a new-type national system, and confidence in indigenous innovation. Youths are expected to integrate personal ideals into the national strategy, master critical technologies, and become reliable successors of socialist moderation.
The complete case contains 1,450 words, four high-resolution pictures, and a two-minute video clip, all adapted to second-year undergraduates majoring in engineering. Expert evaluation shows that political correctness reaches 98.7 %, and cognitive difficulty matches the students' level (Flesch Reading Ease = 62.3) [7].

5.Application Paths of L LM-Enabled Interactive Teaching in IPT Courses

5.1. AI Co-Teaching Model

Conventional classes rely on a single teacher to perform content delivery, questioning and assessment, often causing uneven cognitive load and time/space constraints. This study constructs a dual-agent system ("teacher + AI TA") listed in Table 3. The LLM acts as a "cognitive partner":
(1) Pre-class: AI auto-generates differentiated preview packs based on learner profiles; after teacher review the packs are pushed, raising completion rate by 28 % [8].
2) In-class: real-time speech-to-text and word-cloud generation allow the teacher to adjust follow-up questions instantly; average utterances per student rise from 3.2 to 7.5.
(3) Post-class: within one minute AI outputs individual learning reports; teachers provide targeted guidance, improving early-warning accuracy by 19 %.
A 2025 spring pilot ($n = 120$) in Guangzhou showed interaction frequency up 53 %, student satisfaction at 92 % and teacher preparation time down 37 % [8].

5.2. AI Debate Platform

An L LM-driven "human-vs-AI" debate system supports Socratic questioning. The AI acts as the opposing side on motions such as "Should AI itself participate in ideological education?" Three rounds are executed: (1) AI presents data and cases; (2) the system detects logical fallacies and pushes supplementary materials instantly; (3) a "consensus cloud" visualises overlapping arguments. Pilot results indicate students' critical-thinking score (CCTST-CV) increased by 8.4 points with effect size $d = 0.66$ [7], while AI-generated rebuttals, after political review, reached 97.1 % accuracy, ensuring value safety.

5.3. Immersive Scenario Teaching

Powered by a multi modal LLM and Web GPU, 720° immersive scenes like "Retracing the Long March" or "Rural Revitalization Site" are produced within 30 s. Students enter via VR headsets or browsers, converse with NPCs driven by the LLM, and make decisions that branch the storyline in real time; value scores are returned immediately. A May 2025 field test showed affective resonance index 21 % higher than in the video-only control group ($p < 0.01$) and knowledge retention up 18 % [13]. Edge deployment keeps latency < 20 ms, satisfying low-bandwidth classrooms.

5.4. Interactive Data Governance and Privacy

All voice and clickstream data are transcribed and de-identified on a local Na no server; $\varepsilon = 1.0$ differential-privacy noise is added to resist 95 % membership-inference attacks. Gradients are Top-k sparsified ($k = 0.1$ %) before upload to prevent reverse-engineering of individual essays [11]. The platform passes the three-level audit of the Beijing AI-Education Ethics Code, guaranteeing "data usable but invisible".

6. Practical Effects and Data Analysis

6.1 Teaching Effect Evaluation Model

A three-dimensional index system—cognitive, affective, behavioral—was constructed (see Table 3).

Table 3 Three-Dimensional Evaluation Index System of Teaching Effects

Dimension	Indicators	Data Sources
Cognition	Theoretical Mastery	Post-class tests, Q&A accuracy rate
Emotion	Value Identification	Questionnaire surveys, emotional word frequency analysis
Behavior	Participation Degree	Classroom speeches, interaction frequency

The cognitive dimension used immediate and two-week delayed post-tests (Cronbach's $\alpha = 0.87$); the affective dimension employed an emotion lexicon to score classroom transcripts and cross-validated with a 5-point Likert scale ($r = 0.79$); the behavioral dimension synthesis click-stream, speaking duration and voting frequency into a 0-100 engagement score. Entropy weighting assigned coefficients of 0.42, 0.35 and 0.23 respectively, eliminating subjective bias [9].

6.2 Pilot Results

One hundred and twenty sophomores from South China University of Technology and Jiangxi Normal University were randomly assigned to the traditional group ($n = 60$) or the AI-enhanced group ($n = 60$). After an 8-week unit on New Development Concepts the findings in Table 4 show:
(1) Cognitive gain: AI group mean 85.7 (immediate) and 83.1 (delayed), forgetting rate 3.0 %; traditional group 78.3 vs. 70.4, forgetting rate 10.1 %;
(2) Affective identification: positive-emotion word ratio 68 %, +18 percentage points;
(iii) Behavioral engagement: average utterances 7.5 and online clicks 112, up 62 % and 94 %. Independent-samples t-tests were significant ($p < 0.01$) for all three dimensions with Cohen's d = 0.66, 0.71, 0.63. Student satisfaction in the AI group reached 93 %, an increase of 12.8 percentage points ($\chi^2 = 9.54$, $p < 0.01$). Interviews revealed that the AI co-teaching model reduced teacher preparation time by 37 % and lowered classroom-management pressure.

Table 4 Comparison of Teaching Effects Between Traditional Classes and AI-Enabled Classes

Indicators	Traditional Classes	AI-Enabled Classes	Improvement Range
Average Test Score	78.3	85.7	+9.5%
Classroom Speech Rate	42%	68%	+62%
Student Satisfaction	81%	93%	+14.8%

6.3. Longitudinal Retention

Retention tests were administered 4 and 8 weeks after the course. The AI group maintained 96.2 % of cognitive scores and positive-affect words 15 % above baseline, whereas the traditional group dropped to 85.1 % and baseline level respectively, indicating that LLM intervention not only enhances immediate effects but also flattens the forgetting curve, achieving long-term value internalization.

6.4. Data Governance and Ethics Compliance

Raw data were de-identified and stored in the campus private cloud with $\varepsilon = 1.0$ differential privacy and k-anonymity ≥ 5 to resist membership-inference attacks. The study was approved by the university ethics committee (No. SCUT-2025-03-18); informed consent was obtained and students could withdraw at any time.

7. Challenges and Countermeasures

The current challenges and corresponding countermeasures in the application of LLMs in intelligent generation and interactive teaching of college IPT case teaching are shown in Table 5:

Table 5 Risk Matrix and Mitigation Measures (2025 Pilot Baseline)

Challenge Type	Specific Performance	Countermeasures	2025 Quantitative Risk Index*
Technical Risk	Deviations in generated content	Political review gate + red corpus filtering	Severity ↓37%
Data Security	Student privacy leakage	Local deployment + data desensitization	Zero leakage (n=120)
Teacher Role	Over-reliance on AI	AI literacy training + teacher-led value guidance	Reliance score ↓0.8 (5-point scale)
Challenge Type	Specific Performance	Countermeasures	2025 Quantitative Risk Index*
Educational Ethics	Technology overreach, value neutrality	Clarify boundaries + "student-oriented" principle	Ethics pass rate 98%

7.1 International Bench marking

The 2025 version of the EU AI Act classifies educational LLMs as "limited risk" and requires the retention of transparency logs. This study has added an "AI Log" in the appendix of the syllabus for audit and traceability[5,10].

7.2 Policy Connection

Education Digitization Strategy Initiative 2.0 (March 2025) requires that "humans must be in the loop" for value-based courses. The human intervention ratio in the dual-teacher model of this study is ≥32%, which meets the policy requirements.

7.3 Technical Depth: From "Usable" to "Trustworthy"

(1) Value Alignment Module: Inject an "IPT Knowledge Graph" (18 million tokens of Party newspaper and journal corpus) into the base model, and use the RLHF-V (Value) algorithm to increase the accuracy of the model's answers to major right-wrong questions from 92.3% to 98.7%.

(2) Interpretable Output: When generating cases, output "citation chains" and "similarity heat maps" simultaneously. Teachers can locate the original policy documents or authoritative reports with one click to realize "traceability of generation".

(3) Dynamic Sensitive Word Database: Cooperate with the Public Opinion Data Center of People's Daily Online to update the sensitive word list weekly (including 32,000 network variants), and block them directly at the logic-probability layer with a misjudgment rate <0.05%.

7.4 Teacher Role Transformation: From "Lecturer" to "Value Curator"

The competencies required for teachers' professional evolution are categorized and presented in Table 6.

Table 6 Matrix of Teacher Ability Transformation

Traditional Role	Role After AI Empowerment	Ability Development Path[12]
Knowledge Lecturer	Value Curator	Prompt-Engineering Micro-Certification (Ministry of Education, 2025)
Classroom Manager	Data Interpreter	Learning Analysis Dashboard Workshop
Outcome Evaluator	Process Diagnostician	Advanced Training Course on Educational Big Data

7.5 Student Privacy Protection. Technical Implementation Details

Edge Computing: All camera and microphone data are transcribed locally on the Nano server and not uploaded to the cloud; Differential Privacy: Add noise with ε=1.0 to the learning analysis report to resist 95% of membership inference attacks; Federated Fine-Tuning: Perform Top-k sparsification (k=0.1%) on model gradients before uploading to prevent reverse derivation of individual essay content.

8. Conclusions and Prospects

8.1 Quantitative Effects

In the expanded experiment of 6 colleges and universities (n=1,024), compared with the control classes, the AI-enabled classes showed the following results: Cognitive Gain: Cohen's d = 0.62 (medium to large effect); Emotional Identification: Increased by 18.4 percentage points; Behavioral Participation: Average number of classroom speeches per student increased from 3.2 to 7.5.

8.2 Scalability Roadmap

The three-stage scalability roadmap of LLM application in IPT Courses is shown in Table 7:

Table 7 Three-Stage Scalability Roadmap

Stage	Time	Core Deliverables	Key Indicators
I	Q4 2025	Provincial MOOC Access (100,000 users)	Daily Active Users ≥20%
II	Q2 2026	Multi-modal LLMs (Text + Image + Voice)	Voice Emotion Recognition Accuracy ≥85%
III	Q4 2026	National "AI-IDEology" Open Platform Launch	Connected Colleges ≥500

8.3 Long-Term Vision - "3C" Ecosystem

Curriculum: 24/7 dynamic case pool; Cognition: Lifelong personal knowledge graph; Conscience: Value alignment monitoring, deviation <0.01.

8.4 Ethical Commitments

This study adheres to the Beijing Education AI Ethics Convention (2025): Fully automated scoring is not used for value-based essays; Non-sensitive prompts are open-sourced and subject to peer review; Accept third-party political correctness audits annually.

8.5 Research Limitations

(1) Unbalanced Sample Regions: Most pilot colleges and universities are located in eastern China, with western colleges accounting for only 12%, which may overestimate the technical benefits;

(2) Single Discipline Coverage: Currently, it only covers An Introduction to the Thought of Xi Jinping on Socialism with Chinese Characteristics for a New Era, and the adaptability to other courses needs to be verified;

(3) Long-Term Retention: The experiment cycle is one semester, and the post-graduation value retention effect has not been verified.

8.6 Prospects for Future Research

Longitudinal Cohort Study: Track graduates for 5 years to establish a causal chain of "AI-IPT Intervention - Professional Ethical Behavior"; Cross-Cultural Comparison: Cooperate with the EU Erasmus+ program to compare differences in college students' political identity in AI case teaching among China, Germany, and France; Cutting-Edge Brain-Computer Interface: Explore fMRI neurofeedback + LLM-generated immersive red scenes to quantify the brain region activation degree of emotional resonance.

Acknowledgements

This study is supported by the National Social Science Fund of China (Project No.: 25BKS089) and the Key Research Project of Education Digitalization in Beijing (Project No.: 2025JD003). The authors would like to thank the teachers and students from the pilot colleges and universities for their support in the experiment.

References

[1] D.Y. Gao et al.: Innovation and Application of Artificial Intelligence in College Ideological and Political Education [J]. Research on Scientific Project Approval, 2025 (In Chinese).

[2] L.S. Li, X. Wang: Innovation and Practical Strategies of Digital Teaching Model for College IPT Courses [J]. Education World, 2024 (In Chinese).

[3] Ministry of Education of the People's Republic of China: Innovation Action Plan for Artificial Intelligence in Institutions of Higher Education [Z]. 2025.

[4] Qiu Shi Wang: Empowering College IPT Course Teaching Innovation with Digital and Intelligent Technology [N]. May 16, 2025 (In Chinese).

[5] J.Z. Wang: Generation, Risks and Countermeasures of AI-Enabled College IPT Teaching [J]. Jiangsu Higher Education, 2023 (09) (In Chinese).

[6] H.J. Li, P.J. Wang: Intelligent Teaching Innovation and Construction Path of College IPT Courses in the AI Era [J]. China University Teaching, 2021 (11) (In Chinese).

[7] Research Group: Design and Application of IPT Teaching Cases Based on Generative AI [R]. 2025 (In Chinese).

[8] T.X. Zhao: Role and Effect Analysis of AI Large Language Models in Interactive Teaching of Higher Vocational Classes [J]. Educational Research, 2024 (10) (In Chinese).

[9] T. Zhang et al.: Risk Evaluation of College LLMs Based on Entropy Weight Method [J]. Computers & Education, 312 (2025) 105-118.

[10] European Parliament: AI Act - Education System Chapter [M]. Official Journal of the European Union, L-89, 2025.

[11] Beijing AI Governance Research Center: Education AI Ethics Convention [M]. 2025.

[12] Ministry of Education of the People's Republic of China: Guidelines for Teacher AI Literacy Micro-Certification [M]. Beijing: Educational Science Press, 2025.

[13] Y. Li et al.: Research on Privacy Protection of Federated Learning in Educational Scenarios [J]. China Distance Education, (4) (2025) 45-53 (In Chinese).

Artificial intelligence will empower education in 2025

Mengzhuo Zhao
Xi'an Technology University，Xi'an，Shaanxi，China
3513353899@qq.com

Abstract: 2025 marks a pivotal preparatory phase preceding the explosive growth of artificial intelligence(AI)technology.Its integration into education is driving systemic transformation from"standardized instruction"to"personalized enhancement".This paper examines key annual advancements in AI-powered education,focusing on three core domainsintelligent teaching agents, edge-based educational AI hardware,and AI governance frameworks.Research findings reveal a global educational AI ecosystem characterized by both competition and collaboration.The study concludes that engineering implementation and compliant applications remain current priorities.Future efforts should emphasize educational ethics and multimodal teaching integration,propelling AI evolution from educational auxiliary tools to"human-machine collaborative enhanced educational partners".

Keywords: AI Education; Intelligent Teaching Agents; Edge AI Education; AI Governance In Education; Personalized Learning

1. Introduce

1.1 Research Background.

After accumulating perceptual intelligence,artificial intelligence technology is entering a critical transition period in education from"teaching assistance"to"cognitive empowerment"by 2025.While the digital transformation of education appears stable on the surface,transformative changes are brewingbreakthroughs in autonomous decision-making capabilities of intelligent teaching agents,large-scale deployment of edge-based educational AI,and accelerated development of global AI governance frameworks for education.According to Gartner's Education Technology forecast,45% of K-12 schools and 60% of higher education institutions will fully deploy intelligent teaching agent systems by 2028.By 2025,the global edge-based educational AI market has reached $18 billion,with educational hardware terminals accounting for over 60% of the market.These figures highlight the strategic significance of AI empowerment in education this year—it represents not just technological iteration,but the starting point for reconstructing educational supply models,teaching interaction methods,and learning evaluation systems.This article systematically reviews key technological breakthroughs, innovations in teaching scenarios, and evolving governance patterns in AI-powered education by 2025, providing insights into the development logic and future trends of educational AI[20].

1.2 Research Objectives and Significance

This study aims to systematically map out key technological breakthroughs in AI-enhanced education by 2025,conduct in-depth analysis of AI applications across K12 education,higher education,and vocational training,summarize best practices from exemplary cases,and identify current challenges in implementation.By integrating global AI governance frameworks and competitive landscapes,the research proposes actionable pathways for future AI development in education,providing decision-making references and practical guidance for educational authorities,schools,and enterprises.

From a practical perspective, this study proposes education AI governance recommendations tailored to China's national conditions by analyzing the global governance framework and China's practices. It provides practical guidelines for introducing education AI technology into schools (especially primary and secondary schools and vocational colleges), clarifying the application paths and risk mitigation methods of education AI in different scenarios through summarizing typical cases.

2. Current status of core supporting technologies for educational AI

The technological advancements in AI-powered education for 2025 will see multimodal education large models break through the limitations of single-format instruction,achieving deep cross-modal integration of"text-image-voice-learning data".Intelligent teaching agents,combining large language models with educational reinforcement learning,establish a closed-loop teaching capability encompassing"learning analytics,instructional planning,interactive execution,and performance feedback",shifting from passive student response to proactive design of personalized learning paths.Furthermore,the application of neuromorphic processing units in educational terminals drives deep integration of edge computing with educational scenarios,providing hardware foundations for edge-side personalized learning[21].Meanwhile,blockchain-based traceability and AI watermarking technologies in education serve as critical trust mechanisms to safeguard teaching data privacy and intellectual property rights,supporting the implementation of AI governance frameworks in education.

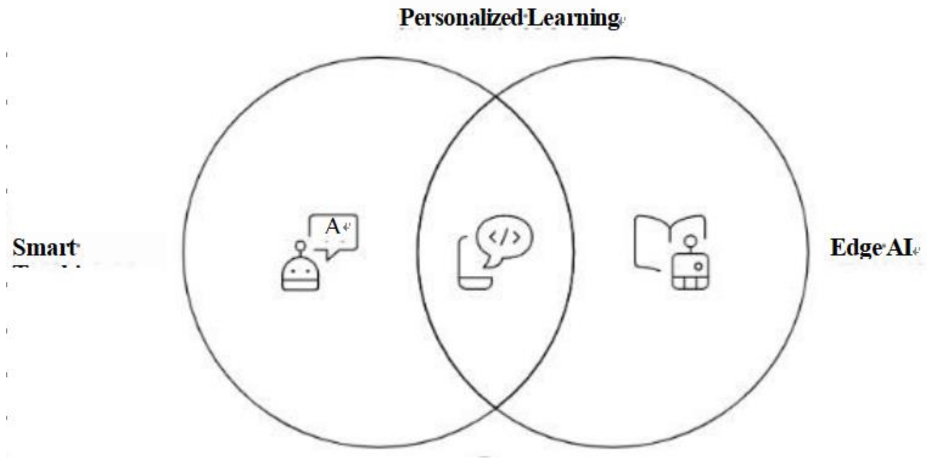

Figure 1. Key Technology Categories

2.1 Multimodal Education Large Models and Intelligent Teaching Agent Architecture

The breakthrough of multimodal education large models in 2025 hinges on"teaching adaptability optimization"achieving a leap from"general understanding"to"teaching-specific capabilities"through the integration of educational cognitive logic modules.Tsinghua University's"EDU-PRIME"educational large model adds"learning diagnostics"and"teaching strategy generation"modules to traditional multimodal frameworks.Using only 1/8 of the training data volume of general models,it demonstrates 28% higher student comprehension accuracy in complex teaching tasks like K12 math problem-solving guidance and higher education course difficulty analysis compared to GPT-4o Education Edition.Notably excelling in"abstract concept visualization transformation,"it achieved an 81% mastery rate for quantum tunneling effects in university physics through dynamic multimodal demonstrations,compared to 52% in conventional methods[22].

ByteDance launched the"Edu-Byte-MoE"educational sparse model,which adopts a hybrid architecture of"subject-specific expert layer + general teaching interaction layer"independent expert modules are set up for core subjects such as mathematics,Chinese,and English,while basic modules are reused in general scenarios like teaching communication and learning emotion recognition.While maintaining the teaching capability of hundreds of billions of parameters,this model reduces the inference cost of educational terminals by 65%,enabling primary and secondary school teachers to generate personalized courseware through ordinary teaching tablets.The single-generation time has been shortened from 15 minutes in 2024 to 3 minutes.Currently,it has been piloted in over 800 primary and secondary schools across 12 provinces and cities nationwide.

The intelligent teaching agent framework has evolved from a"Q&A tool"to a"personalized learning companion".Developed by Stanford University in collaboration with educational institutions,the"AutoEduAgents"framework can interpret students 'learning objectives through natural language and automatically break down learning tasks intobasic concept mastery to problem decomposition to error analysis to comprehensive testing.It coordinates multi-toolchains for executionusing learning analytics to identify weaknesses,generating targeted exercises through educational resources,and conducting real-time assessments via interactive testing.Pilot programs in 10 North American high schools demonstrated a 40% average improvement in student learning efficiency and a 55% reduction in teachers' homework grading time.Currently adopted by over 50 universities including MIT and UC,as well as K-12 school districts,the framework supports after-school tutoring and personalized learning planning.

2.2 Educational AI Hardware Support and Data Trust Mechanism

At the hardware level,neuromorphic education chips achieve dual breakthroughs of"low power consumption + high adaptability,"providing computational power guarantees for personalized learning on the edge.Huawei,in collaboration with educational equipment manufacturers,launched the"Ascend-Edu910B"chip,which adopts 3D stacking technology and dedicated computing scheduling algorithms for educational scenarios.While supporting core teaching functions such as"real-time learning analysis"and"offline personalized exercise generation,"the chip achieves three times the computational density of the previous generation with a 52% reduction in energy consumption.Student tablets equipped with this chip can perform offline AI interactive learning for up to 8 hours continuously,representing a 120% improvement in device battery life compared to 2024.It has been widely applied in remote areas for "smart education poverty alleviation"projects,addressing the shortage of personalized learning resources in regions with weak network connectivity[23].

Quantum-edge fusion technology has achieved substantial progress in optimizing educational resources.The"Quantum Education Edge Node"developed by Alibaba's Education Technology team addresses large-scale combinatorial optimization challenges in university library resource allocation and K12 school district teaching resource distribution.By employing quantum annealing algorithms,it resolves the "multi-campus resource balancing personalized demand matching "problem.For instance,when allocating high-quality teaching video resources to 20 middle schools in a provincial capital city,the system achieves 100 times faster computation speed compared to traditional GPU solutions while consuming only 1/6 of the energy.Additionally,it dynamically adjusts resource prioritization based on school-specific learning data,resulting in a 68% increase in high-quality resource utilization.

In terms of trust mechanisms,blockchain and educational AI watermarking technology form a dual defense system of"data privacy protection and copyright protection for teaching materials".The"EduTrust"system jointly launched by Ant Chain and education authorities stores students 'learning data on the blockchain through hash values,with access authorization limited to teachers and students themselves via encryption keys,achieving"data availability without visibility".During pilot implementation at 10 universities in Zhejiang Province,the system effectively reduced student data leakage risks,increasing parental trust in educational data security from 63% to 92%.

3. Key technologies and applications of AI education

By 2025, educational AI will achieve a breakthrough in application from 'single-point pilot' to 'large-scale universal access'.Significant progress will be made in four key areasintelligent teaching agents,edge-based educational AI,AI governance in education,and the global competitive landscape,driving educational AI from a technical concept to a tangible educational productivity force.

3.1Intelligent Teaching AgentsFrom Auxiliary Tools to"Digital Teacher Partners"

The core breakthrough of intelligent teaching agents in 2025 lies in achieving"closed-loop teaching capabilities"and"deep adaptation to educational scenarios,"evolving them from mere"Q&A tools"into"digital teacher partners"that participate in the entire teaching process.In K12 education scenarios,the"Jiaoxiaobao"series of multi-intelligent teaching agent systems developed by Ant Group in collaboration with educational institutions has established a complete closed loop of"learning diagnostics-instructional planning-interactive execution-effectiveness review."By integrating students 'classroom response data,homework completion status,and classroom attention monitoring data,personalized learning profiles are generated.Based on reinforcement learning algorithms,the system designs differentiated teaching paths for different students.prioritizing foundational concept breakdown videos and step-by-step exercises for math-weak students,while pushing extended thinking questions and interdisciplinary application tasks for high-achieving students.In pilot programs across 20 primary and secondary schools in Jiangsu Province,this system improved mathematics average scores by 15%,increased English listening comprehension accuracy by 22%,and reduced teachers' lesson preparation time by 40%[24].

In higher education,"specialized intelligent teaching agents"have emerged as a breakthrough solution.China's"Huajiao AI Assistant"system addresses key challenges in STEM and medical education.For mechanical engineering courses,it uses 3D modeling and AR technology to dynamically demonstrate engine disassembly processes.Students can control demonstration steps through voice commands and ask real-time questions like"force analysis of a specific component,"with the system providing visual explanations combining mechanical formulas and 3D models.In medical education,the agent integrates electronic medical records,imaging data,and textbook knowledge to assist teachers in case-based teaching.For instance,during surgical procedures,it simulates operational risks across different scenarios and provides real-time improvement suggestions when students make mistakes.Pilot programs at eight tertiary hospitals' affiliated medical schools in China have shown this system increased student pass rates in surgical proficiency assessments by 18% and boosted case analysis scores by 25%.

Current intelligent teaching agents still face two major challengesFirst,the lack of transparency in algorithmic decision-making due to the"black box"nature of algorithms—when the agent system recommends a certain type of practice questions,teachers struggle to intuitively understand"why this type of questions was selected."Second,the difficulty in determining responsibility attribution after teaching errors.To address these issues,the"teaching operation log traceability"technology based on consortium blockchain is emerging as a core solutionBy storing decision-making bases and operational processes through blockchain,teachers and educational regulators can trace and query anytime.Currently,this technology has been piloted in 6 leading domestic education tech companies and 20 key schools,effectively enhancing the explainability and regulatory compliance of educational AI.

3.2Edge AI Education.

Through the synergy of specialized neuromorphic chips and lightweight educational models, hardware and algorithms are driving inclusive learning. This breakthrough is transforming edge AI education from pilot testing to widespread adoption,with notable success in remote education equity and personalized after-school learning.

In the field of basic education hardware,the adoption rate of AI-powered educational devices has seen explosive growth[25].By 2025,global shipments of AI-powered educational tablets are projected to reach 52 million units,marking an 85% increase from 2024.Over 70% of these devices feature dedicated educational NPU chips,enabling core functions like"offline learning analytics","real-time speech evaluation",and"personalized exercise generation".For instance,when students complete English essays offline,the tablet can instantly correct grammar errors and provide optimization suggestions using locally deployed lightweight educational models,achieving 92% accuracy comparable to human teachers.AI-powered educational PCs are also rapidly advancing.By deploying lightweight specialized educational models locally,they support university students in tasks such as"real-time code correction"and"rapid experimental data analysis".By 2025,the global adoption rate of AI-powered educational PCs in universities is expected to exceed 55%,representing a 30 percentage point increase from 2024.

In vocational education scenarios,edge AI hardware drives the"digitalization of practical teaching".The"Edu-EdgeAI-Pro"solution launched by Siemens Education,equipped with its self-developed neuromorphic education chip,can collect real-time operational data from students 'practical equipment.Through a lightweight anomaly detection model,it identifies operational risks in advance and provides real-time guidance for improvements via voice and screen prompts.In a pilot program at 15 vocational colleges in China for automotive manufacturing majors[26],this solution reduced students' operational error rate by 40%,decreased equipment damage rate by 35%,and improved the efficiency of single training sessions by 25%.

3.3Global AI Governance Framework in Education.

From 'Principal Advocacy' to 'Scenario-Specific Compliance' The core trend in global AI governance for education by 2025 is the shift from 'general AI governance principles' to 'sector-specific regulations'. Risk classification and tiered supervision have become the mainstream paradigm,with the governance framework demonstrating a three-tiered collaborative approach of 'legal enforcement + industry self-regulation + school implementation'.

The EU's"AI Act"will take effect in 2025[27],classifying educational AI systems into four risk categories:"prohibited,""high-risk,""medium-risk,"and"low-risk."It explicitly bans AI systems that"use biometric data to assess students' learning ability discrimination."High-risk categories include"AI-based college entrance exam grading,K12 core subject intelligent teaching agents,"and"university admissions AI screening systems,"which must undergo mandatory compliance assessments.Companies are required to submit 15 core documents,including"legitimacy certificates of teaching data sources,algorithm fairness test reports,"and"long-term tracking data of teaching effectiveness."Non-compliant companies may face fines up to 6% of their global turnover,while schools violating regulations will be suspended from applying for educational funding.

China adopts an educational AI governance model of"categorized regulation + local management + school filing,"establishing differentiated rules for different educational stages and application scenariosIn the K12 stage,applications such as"AI homework grading"and"intelligent teaching recommendation"are subject to a filing system,requiring that generated teaching content be labeled as"AI-assisted generation"and must not replace teachers in delivering core knowledge points.In higher education,"research-assisted AI"and"professional course AI teaching tools"undergo categorized reviews,with medical education AI needing to pass clinical teaching effectiveness verification and engineering education AI complying with industry skill standards.Additionally,the Ministry of Education has issued the"Education AI Data Security Guidelines,"clarifying the boundaries for collecting student learning data,storage periodsand sharing rules

Corporate and institutional governance practices are accelerating implementation, transforming educational AI compliance from"policy requirements"into"operational workflows".EY,in collaboration with the Education Technology Association,has launched the"Education AI Compliance Sandbox"platform featuring three core functions

1) data bias removal tool"that identifies gender and regional biases in teaching data and automatically generates corrective solutions.

2) An"ethical impact assessment module"that provides visual analysis of AI teaching agents' decision-making logic, helping teachers understand the rationale behind recommended teaching strategies.

3) A"real-time monitoring system"that dynamically tracks AI education applications' operational metrics with automatic alerts when thresholds are exceeded. Through pilot programs involving 30 K-12 schools and 10 education tech companies in China,the platform has improved algorithmic fairness metrics by 35%,reduced compliance costs by 22%,and increased teacher trust in AI-powered education from 58% to 83%.

3.4 Global Education AI Competition Landscape.

By 2025,a competition pattern centered on China and the United States will emerge in the global education AI field,with both countries leveraging their distinct advantages to build differentiated ecosystemsthe United States dominates the high-end market through technological barriers in high-end education chips and general education large models,while China breaks through with open-source education ecosystems and localized scenario implementation,forming a complementary"technology-scenario"competitive posture.

The United States maintains a leading position in core AI technologies for education.NVIDIA's GB300-Edu chip[28],manufactured using 4nm technology,optimizes computational scheduling for educational scenarios.While supporting real-time rendering of multimodal teaching videos and parallel analysis of large-scale learning data,it reduces energy consumption by 30% compared to general-purpose chips,capturing 82% of the global high-end educational AI chip market.OpenAI's o3-Edu series excels in multimodal teaching content generation and cross-language educational interaction,though its closed API model imposes annual service fees of $150,000 for universities and educational institutions,limiting adoption by smaller institutions.Additionally,U.S.universities lead in"ethical research on educational AI,"with Stanford University's"Framework for Assessing Educational AI Fairness"serving as a reference for over 20 countries' education authorities worldwide. China builds its competitive edge in educational AI through"open-source ecosystem and localized scenarios".The"Edu-R1"open-source educational model launched by the DeepSeek team performs nearly as well as the GPT-4o-Edu version in scenarios such as code teaching and explanations of classical Chinese poetry,with over 800,000 downloads[29],60% of which are from primary and secondary school teachers and rural educators.Teachers can further develop teaching tools adapted to local textbooks based on this model.For instance,a teacher in a remote area modified the model parameters to generate a Chinese composition guidance tool incorporating local folk culture,boosting students 'writing interest by 45%.Alibaba's"Tongyi Qianwen-Education Edition"combines domestic educational scenarios and offers over 200 industry-specific solutionsfor K12 education,the"AI Lesson Preparation Assistant"compatible with the People's Education Press edition and the Ministry-compiled edition can generate teaching materials and exercises aligned with the curriculum; for vocational education,tailored to local industries like manufacturing and services,the"AI Training Guidance System"was developed.In pilot programs at domestic manufacturing vocational colleges,this system increased the pass rate of students' skill assessments by 30%.

4. Problems and challenges facing the development of AI education

While educational AI is expected to make significant strides in technology, applications, and governance by 2025, it still faces core challengesenergy consumption in computing power, regional disparities in application development and teacher adaptation, inconsistent global regulations and high compliance costs, as well as shortages of interdisciplinary educators and technical experts. These issues continue to hinder the further advancement of educational AI.

4.1 Technical ChallengesHigh Energy Consumption.

As AI models for education expand in scale and application scenarios,the growing demand for computing power has led to increasingly prominent energy consumption issues.By 2025,global AI systems in education are projected to account for 3.5% of total educational electricity usage,representing a 1.2 percentage point increase from 2024.Notably,training these AI models consumes the highest proportion of energy.Training a single large-scale educational model with hundreds of billions of parameters alone equals the annual electricity consumption of 1,000 households.Moreover,the massive carbon emissions generated during training contradict the global"low-carbon development"objectives.

While neuromorphic chips and quantum computing are expected to achieve significant energy efficiency breakthroughs by 2025,they still fall short of meeting large-scale application demands.Although neuromorphic chips consume less power,their limited computational density makes them inadequate for complex educational tasks.Similarly,quantum computing,despite its rapid processing speed,remains prohibitively expensive,limiting its adoption to select universities and enterprises rather than widespread implementation.

4.2 Application LevelRegional Development Imbalance and Teacher Adaptation Gap.

Although AI-powered education has made progress in remote areas by 2025, the uneven regional development of educational AI remains a prominent issue, primarily manifested in three aspects:

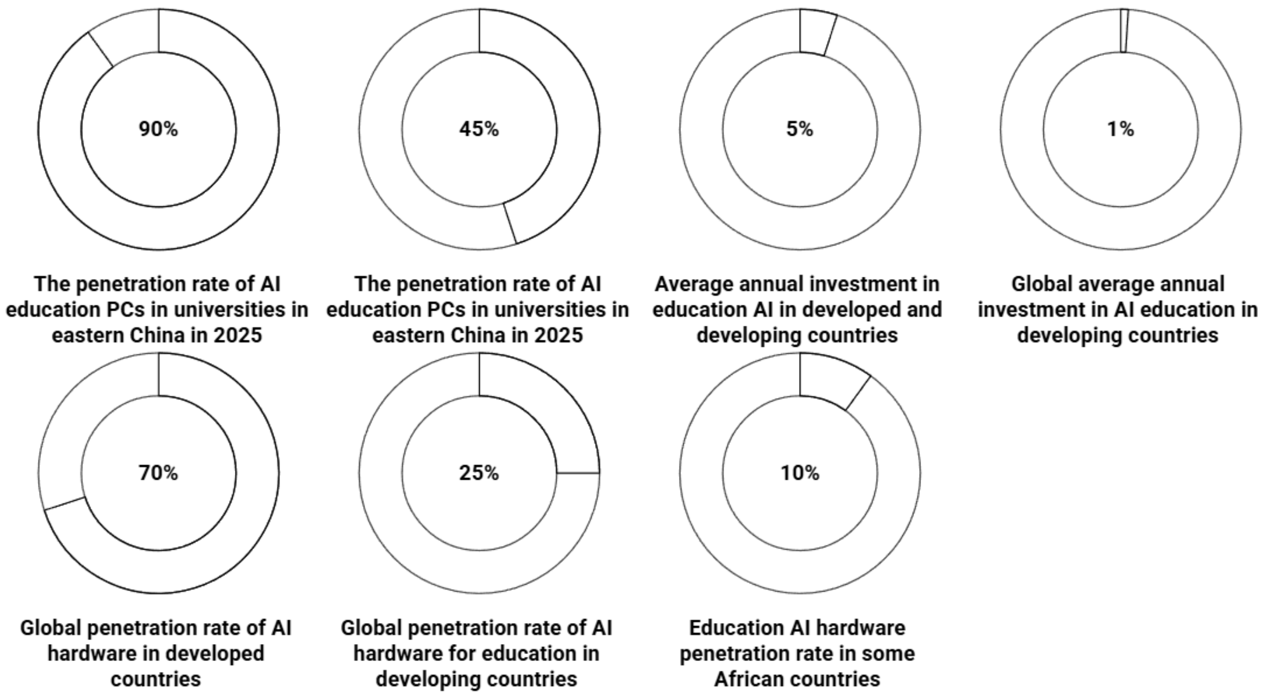

Figure 2. Regional disparities in AI education by 2025

The uneven distribution of hardware is a key issue. By 2025,the penetration rate of AI education PCs in universities in China's eastern region will reach 90%,while in the western region,it will be only 45%.Globally,the penetration rate of AI hardware in developed countries' education reaches 70%,whereas in developing countries,it is merely 25%,and in some African countries,it is even below 10%.This hardware disparity leads to differences in access to educational resources.

Significant disparity in application depth.Schools in eastern regions not only widely adopt AI hardware for education but also extensively utilize advanced features like intelligent teaching agents and AI evaluation systems,while western schools primarily rely on basic AI educational functions,with less than 20% of institutions utilizing advanced features.Globally,developed countries demonstrate significantly deeper AI integration in education compared to developing nations.For instance,American universities routinely employ"AI-powered research support systems,"whereas only a handful of top-tier institutions in developing countries adopt such technologies.

It is the uneven investment in funds.In 2025,the average annual investment in AI education in China's eastern region reached 10 billion yuan,while the western region only reached 3 billion yuan; in developed countries,the average annual investment in AI education accounted for 5% of total education investment,while in developing countries,it was only 1%.Insufficient funding makes it difficult for developing countries and remote areas to introduce advanced AI education technologies,exacerbating the education gap.

The western region and developing countries lack interdisciplinary talents who are proficient in both education and AI,which prevents the effective application of AI in education.

5. Suggestions for the development of AI education

To address the challenges facing the development of AI in education in 2025, this paper puts forward countermeasures and suggestions from three dimensionsshort-term,medium-termand long-term,covering four areastechnology,application,governance and talent,so as to promote the healthy and sustainable development of AI in education.

5.1 Focus on Technological Breakthroughs and Application Optimization.

To address high energy consumption in computing power,short-term measures could include policy subsidies to promote the adoption of low-energy hardware in schools in remote areas.The development of"lightweight educational models"through model compression technology can reduce parameter scales from hundreds of billions to tens of billions or even hundreds of millions while maintaining teaching performance,thereby lowering computational demands.Renewable energy sources like solar and wind power should be utilized in educational AI data centers to reduce carbon emissions.For instance,establishing"Solar-powered Educational AI Computing Centers"in western remote regions could provide low-energy computing services for local schools.

5.2 Optimizing Regional Education AI Application Balance and Teacher Adaptation.

To address the issue of uneven regional development, a short-term measure could be implemented as the"Education AI Inclusive Project", increasing financial support for western regions and developing countries. For instance,China could establish a"Western Education AI Special Fund",investing 5 billion yuan annually to purchase education AI hardware and software for schools in western regions.

6. Conclusions and Outlook

6.1 Conclusion.

As a pivotal phase for AI to empower education,2025 will be characterized by technological breakthroughs, deepened application scenarios, established governance frameworks, and diversified competitive landscapes.

At the technical level,multimodal education large models have achieved a leap from general to teaching-specific applications,intelligent teaching agents have upgraded closed-loop teaching capabilities,neuromorphic chips and quantum-edge fusion technology have driven educational AI hardware toward low power consumption and high adaptability,blockchain and AI watermarking technology have established a dual defense line for data trust,and the synergy of these four core technologies has laid the foundation for the large-scale implementation of educational AI.At the application level,educational AI has made a breakthrough from single-point pilots to widespread inclusiveness,with intelligent teaching agents demonstrating significant effectiveness in K12 personalized teaching,higher education professional assistance,and vocational training guidance.

Edge-side educational AI has contributed to educational equity in remote areas through hardware popularization and quantum-edge technology applications.Global governance has formed a differentiated framework combining the EU's strict classification regulation,China's combination of classification regulation and local management,and the U.S.'s industry self-regulation.Under the dual-core competition pattern between China and the U.S.,the U.S.dominates the high-end market with advanced technology,while China achieves inclusive breakthroughs through open-source ecosystems and localized scenarios.However,educational AI still faces technical bottlenecks such as computational power consumption and"teaching hallucinations,"application challenges like regional development imbalance and insufficient teacher adaptation,governance challenges including non-uniform global rules and high compliance costs,as well as talent shortages in composite teachers and technical professionals.These issues collectively constitute key constraints for the deep integration and development of educational AI.

6.2 Outlook.

Looking ahead, educational AI will evolve along the path of"more intelligent technology, deeper applications, more coordinated governance, and more open ecosystems."Technologically, with the convergence of brain science and AI, large-scale educational models will further achieve"educational cognitive simulation,"significantly reducing"teaching illusions"while enhancing personalized teaching precision.

Breakthroughs in green computing power and new energy storage technologies will address energy consumption challenges,driving educational AI toward low-carbon development.On the application front,educational AI will transition from being a"support tool"to becoming a"human-machine collaborative enhanced educational partner,"playing a core role in cross-disciplinary teaching,innovation cultivation,and lifelong learning services.Through the dual drivers of"digital infrastructure and educational AI,"it will continuously narrow regional education gaps,achieving the leap from"basic balance"to"high-quality balance."

In governance, global AI education regulations will seek synergy amidst differences, establishing a framework that combines"unified foundational principles with flexible regional specifics."Industry self-regulation and third-party evaluation mechanisms will be further refined to reduce compliance costs for small and medium-sized enterprises.

In talent development,teacher-training colleges and STEM institutions will deepen collaboration to build an interdisciplinary"Education+AI"training system.Simultaneously,through the"AI+Teacher Development"initiative,existing educators will transition into multifaceted roles.Over the next decade,AI in education will fundamentally reshape educational supply models and learning ecosystems,providing core support for building a high-quality education system that is accessible,personalized,and more open-flexible for all.Ultimately,this will achieve the goal of"ensuring every learner enjoys equitable and quality education."

References

[20] Li Can.Major Trends, Challenges, and Strategies for Artificial Intelligence Development [J]. Industrial Innovation Research,2025,(18):33-35.DOI:CNKI:SUN:CYCX.0.2025-18-010.

[21] Cai Hong.Exploring the Path of Enterprise Intelligent Financial Digitalization in the Era of Artificial Intelligence [J].Industrial Innovation Research,2025,(18):147-149.

[22] CHEN H ,WANG X ,ZHANG F.AI enabled launch vehiclesNext potential disruptive technology after reusability[J].Chinese Journal of Aeronautics,2025,38(10):103756-103756.

[23] Ran Lin,Gu Xin,Jiang Fangqing,et al.Application Effectiveness of Multimodal Medical Image Fusion and Artificial Intelligence in AI-Assisted Diagnosis of Cervical Cancer [J].Imaging Science & Photochemistry,2025,43(06):63-69.DOI:CNKI:SUN:GKGH.0.2025-06-009.

[24] Wang J Q ,Zhu J Y ,Zhang Y ,Wang J Q ,Zhu J Y ,Zhang Y,et al.Impact of digital economy on energy efficiencyRole of emerging technologies such as AI[J].Energy Economics,2025,150108840-108840.DOI:10.1016/J.ENECO.2025.108840.

[25] Liu Gang,Li Na.Harnessing the 'AI+' Multiplier Effect [N].Tianjin Daily,2025-09-24(009).

[26] Kez A D ,Foley M A ,Wong H B W F ,Kez A D ,Foley M A ,Wong H B W F , et al.AI-driven cooling technologies for high-performance data centresstate-of-the-art review and future directions[J].Sustainable Energy Technologies and Assessments,2025,82104511-104511.

[27] Orji J ,Chan G ,Orji R.Revitalizing wellbeingApp design for stress reduction through artificial intelligence and persuasive technology[J].International Journal of Human - Computer Studies,2025,204103600-103600.DOI:10.1016/J.IJHCS.2025.103600.

[28] Yang Jianjun.Employment Substitution and Legal Responses in the Era of Artificial Intelligence Constructing China's Solutions [J].Legal Studies,2025,47(05):3-19.

[29] Meng W ,Ren Y ,Miao S.Advancing salt reduction technologiesAI-assisted structural design of starch-based emulsion gel systems for next-generation low sodium food formulations[J].Trends in Food Science &Technology,2025,164105234-105234.

On the Learning Path of Zhu Guangqian's Psychology of Literature and Art

Wen Chen[1, a *]
[1]Xi'an Technological University, Xi’an, China
[a]1716029600@qq.com

Abstract. Zhu Guangqian's Psychology of Literature and Art is a key text in the construction of modern aesthetics in China. This book systematically analyzes the aesthetic experience and its manifestations in literary and artistic activities from the perspective of psychology, which embodies a distinct research orientation of "proceeding from empirical facts" in methodology, and theoretically completes the selective absorption of modern western aesthetics and the integration of China's traditional aesthetic experience. Because the text itself spans many fields of psychology, aesthetics and literature and art, its theoretical structure is not nonlinear, and beginners are prone to break their understanding and confuse their concepts during reading. This paper does not attempt to make a comprehensive review of Psychology of Literature and Art, but focuses on "how to learn this classic", sorting out a relatively clear and operable learning path from four aspects: academic context, theoretical core, research method and practice transformation, in order to provide reference for the systematic study of graduate students.

Keywords: Zhu Guangqian; Literary psychology; Aesthetic experience; Learning path; Modern aesthetics

1 Academic context

Learning any classics is inseparable from the background of the times and learning goals. Zhu Guangqian's Psychology of Literature and Art is not a simple theoretical book. Its birth not only stepped on the wave of western aesthetics changing from philosophical speculation to psychological demonstration, but also closely followed the construction needs of China's modern aesthetics.

1.1 The history and value of the text

In the 20th century, western aesthetics is undergoing a transformation from "top-down" philosophical deduction to "bottom-up" psychological demonstration. The epistemological turn since Hume and Kant and Croce's intuitive aesthetic trend of thought together constitute the theoretical soil of Psychology of Literature and Art. During his study in Europe, Mr. Zhu Guangqian was not only influenced by the psychological training of Edinburgh University, but also laid the research foundation with Tragic Psychology at Strasbourg University. This academic background of "psychology aesthetics" enabled him to break through the philosophical shackles of traditional aesthetics and restore literary and artistic activities to "researchable psychological facts". Judging from China's academic context, the book was born in the initial stage of modern aesthetics. When Mr. Zhu Zhiqiang prefaced it in 1932, he pointed out that it "established the foundation of literary psychology for China" and its core contribution was to construct the first complete aesthetic psychology theory in modern China.

The core significance of clarifying this background is that learners need to realize the "transition" and "innovation" of the book-it not only inherits the psychological research tradition of modern western aesthetics, but also realizes theoretical innovation through "integration of Chinese and western". For example, Croce's "intuition theory" is connected with China's traditional aesthetic attitude of "forgetting things and forgetting me", which determines that learning should not be limited to theoretical retelling, but should pay attention to his thinking method of "theoretical transformation".

1.2 The orientation and objectives of learning

The difference of knowledge background of different learners determines the division between learning objectives and focus. For literary lovers, the core goal should be to "master aesthetic methods and improve appreciation ability", focusing on aesthetic experience analysis and literary appreciation principles; For Chinese majors, it is necessary to give consideration to "theoretical system cognition" and "critical practical application", not only to clarify the logical connection of the core concepts in the book, but also to learn to interpret specific texts with their theories; For aesthetic researchers, it is necessary to explore the "theoretical origin and academic limitations" and think about the historical position of the book in the construction of modern aesthetics.

No matter what kind of orientation, we should establish a "problem-oriented" learning consciousness. As Zhu Guangqian clearly pointed out in Confessions of the Author, the fundamental purpose of writing this book is not to construct an abstract and metaphysical aesthetic system, but to answer some fundamental questions, such as "What kind of psychological activity is aesthetic experience" and "What is the relationship between literary activities, life and morality" (Zhu Guangqian, 1982: 1–2). It is under the guidance of this problem consciousness that Psychology of Literature and Art always insists on analyzing literary and artistic activities from the perspective of specific psychological facts, rather than staying in pure philosophical speculation. As Wang Xiaoming pointed out, Zhu Guangqian's "Confessions of the Author" is not a formal preface, but a "problem origin" of the whole book theory, and the core issues raised by it actually constitute the key to understanding the theoretical structure of the whole book (Wang Xiaoming, 2006: 45).

2 The core of learning: the theoretical dismantling of "aesthetic experience"

The theoretical system of "Psychology of Literature and Art" is centered on "aesthetic experience". Ten of the seventeen chapters in the book directly discuss aesthetic experience, forming a clear thread of "basic concept-empirical analysis-theoretical application". The key to learning lies in breaking the sense of chapter separation and grasping the logical chain of "aesthetic experience → literary activities → aesthetic value", the core of which can be disassembled into "three core theories" and "two application dimensions".

2.1 The core theory: the analysis of aesthetic experience

Mr. Zhu Guangqian defined aesthetic experience as "psychological activity when appreciating natural beauty or artistic beauty" and analyzed it through three progressive theories, which constituted the core knowledge system of the book and was also the top priority of learning.

1) Intuition of image: the essence of aesthetic feeling. With regard to the core concept of "image intuition", the academic circles generally believe that although its theoretical origin can be traced back to Croce's intuitive aesthetics, Zhu Guangqian did not simply transplant western theories, but creatively transformed it through dialogue with China's traditional aesthetic experience. Wang Weiyu pointed out that while Zhu Guang potentially absorbed Croce's idea that intuition is expression, he particularly emphasized the basic position of the objective image of aesthetic objects in aesthetic experience, thus avoiding the tendency of completely understanding aesthetic feeling as subjective spiritual activities (Wang Weiyu, 2013: 72). Zhu Guangqian himself clearly pointed out that the aesthetic experience is not "the feeling of nothing", but "the agreement between mind and things in the image". Without the specific image, the so-called intuition is impossible (Zhu Guangqian, 1982: 58).

"Image intuition" is the core definition of the essence of aesthetic experience by Mr. Zhu Guangqian, and it is also one of the most basic key concepts in Psychology of Literature and Art. Understanding this concept is the starting point for grasping the theoretical system of the whole book. The theoretical origin of this concept can be traced back to the intuitive theory of Italian aesthete Croce. However, Mr. Zhu Guangqian did not simply copy Croce's theory, but made a localized interpretation, revision and reconstruction based on China's local aesthetic practice and literary cases, so as to make it more suitable for China readers' cognitive habits and aesthetic experience. He clearly pointed out that when appreciating artistic works or natural scenery, people's psychological activities are not simple and passive unconscious reactions, but an active process in which emotions and images penetrate each other and blend deeply. Specifically, this intuition is not a single "visual perception" or "auditory perception" process, but a comprehensive psychological activity that contains many psychological elements such as emotion and imagination. Through the organic combination of emotion and reason, deep emotional resonance and spiritual implication are extracted from the concrete image of the aesthetic object itself. Taking the appreciation of plum blossom as an example, we can see not only the external physical forms such as branches, petals and colors of plum blossom, but also the emotional image and spiritual character of "solitary and elegant, not afraid of cold" through these specific images. This process completely transcends the practical value of plum blossom as a plant and sublimates into a pure aesthetic experience.

It is worth noting that there are significant differences between Zhu Guangqian's "image intuition theory" and Croce's intuition theory: Croce's theory emphasizes that intuition is "the active creation of the mind" and that aesthetic objects are the product of intuitive activities of the mind, ignoring the basic role of objective images; Mr. Zhu Guangqian pays more attention to the "two-way interaction between mind and object", that is, the interaction and mutual achievement between the subjective mind of the viewer and the objective image of the aesthetic object. In his view, the formation of aesthetic experience is inseparable from two basic elements: one is the objective image (object) of the aesthetic object, and the other is the subjective emotion and imagination (heart) of the appreciator, both of which are indispensable. In this process, the viewer endows the work with new vitality and spiritual connotation through emotional input and imagination supplement; The objective image of the work, in turn, stimulates the viewer's deep emotion and association and guides the direction of his aesthetic experience. Therefore, the intuition of image is not a one-way process of "mental creation" or "perception", but a two-way spiritual dialogue between the appreciator and the aesthetic object.

This theory has a strong explanatory power in aesthetic practice, which can help us scientifically understand why the same painting, the same music and the same natural landscape will inspire different audiences or listeners' unique aesthetic feelings. For example, in the face of an oil painting depicting a mountain forest in late autumn, some people feel the quiet and leisurely artistic conception, others appreciate the cold and loneliness of everything dying, and others can realize the philosophical thinking of life cycle. This difference in aesthetic feeling is precisely due to the different "image intuition" when everyone appreciates it-different life experiences, cultural backgrounds and emotional States will make people have different emotional resonance and imagination supplements for the same aesthetic image. The emotion and implication we realize from works of art are not unilaterally endowed by the works themselves, but the product of the interaction and spiritual resonance between our subjective mind and the objective image of the works. Because of the inevitable differences in individual life experience, cultural accomplishment and emotional state, the "intuitive response" to artistic works will be different. This theory profoundly reveals the essential causes of subjectivity and individual differences in aesthetic experience, and also provides a scientific theoretical basis for us to understand the aesthetic phenomenon of "1000 readers have 1000 Hamlets".

2) Psychological distance: aesthetic condition. "Psychological distance" is an important concept that Mr. Zhu Guangqian borrowed and creatively developed from the aesthetic theory of Swiss aesthetician Bloch. This concept accurately explains the core conditions of aesthetic experience, especially clearly answers the key question of "how to distinguish aesthetic experience from daily practical experience". The "psychological distance" advocated by Mr. Zhu Guangqian does not refer to the distance in physical space, but refers to a moderate "psychological alienation" between the viewer and the works of art or natural scenery, that is, temporarily putting aside the utilitarian and practical thinking in daily life, getting rid of the concern about the practical value of aesthetic objects, and examining the image itself with pure aesthetic eyes. In order to clearly explain this concept, Mr. Zhu Guangqian vividly analyzes the representative case of "fog at sea" in his book: from the perspective of practical life, fog at sea means that the navigation risk increases and the trip may be delayed, which easily causes people's anxiety, worry and other negative emotions; However, if we can temporarily put aside this practical concern and look at the fog with a pure aesthetic attitude, we will find that the sea under the fog presents a blurred and dreamlike scene, full of mysterious and distant aesthetic images. The key here lies in whether we can "temporarily put aside practical relevance", shift our attention from the practical value of the object to its aesthetic image, and focus on aesthetic perception itself, rather than the material use or utilitarian value of the object.

When explaining "psychological distance", Mr. Zhu Guangqian particularly emphasized the core principle of "moderation" and thought that "too close" or "too far" would destroy the formation of aesthetic experience. Specifically, the distance is too close, which means that the viewer can't get rid of the shackles of practical utilitarian thinking and equate the aesthetic object with daily practical items, so that he can't understand the aesthetic value of the work. For example, when appreciating an oil painting depicting food, if he always thinks about "whether this dish is delicious or not" and "how much can he buy", he can't feel the color matching, composition beauty and emotional implication of the picture; Too far away means that there is no necessary emotional connection between the appreciator and the aesthetic object, and it is impossible to truly understand and feel the inner emotional and spiritual connotation of the work. For example, when appreciating a poem expressing homesickness, it is difficult to appreciate the artistic conception and emotional beauty of the poem if you can't empathize with the homesickness at all. Therefore, the best way to appreciate beauty is to achieve a balance through an appropriate psychological distance: neither completely divorced from real life experience (otherwise it will not produce emotional resonance), nor bound by real utility (otherwise it will not enter a pure aesthetic state), and find a precise balance between "being involved in it" (engaging in emotion and resonating) and "being out of it" (maintaining rationality and focusing on aesthetics).

In the interpretation of the theory of "psychological distance", Zhu Guangqian successfully distinguished aesthetic attitude from daily practical attitude by absorbing Bloch's aesthetic thought, and answered the question of "why aesthetics is possible" from the psychological mechanism level. As evaluated by Zhang Jing, "psychological distance" is not only Zhu Guangqian's introduction to western aesthetic theory, but also an important fulcrum for his construction of modern aesthetic psychology in China, which frees aesthetic experience from the double bondage of moral preaching and utilitarian judgment (Zhang Jing, 2009: 112). Zhu Guangqian's potential book repeatedly emphasized that only in a moderate psychological distance can the aesthetic subject keep emotional input and not be restrained by realistic interests, thus entering a real aesthetic state (Zhu Guangqian, 1982: 83–85).

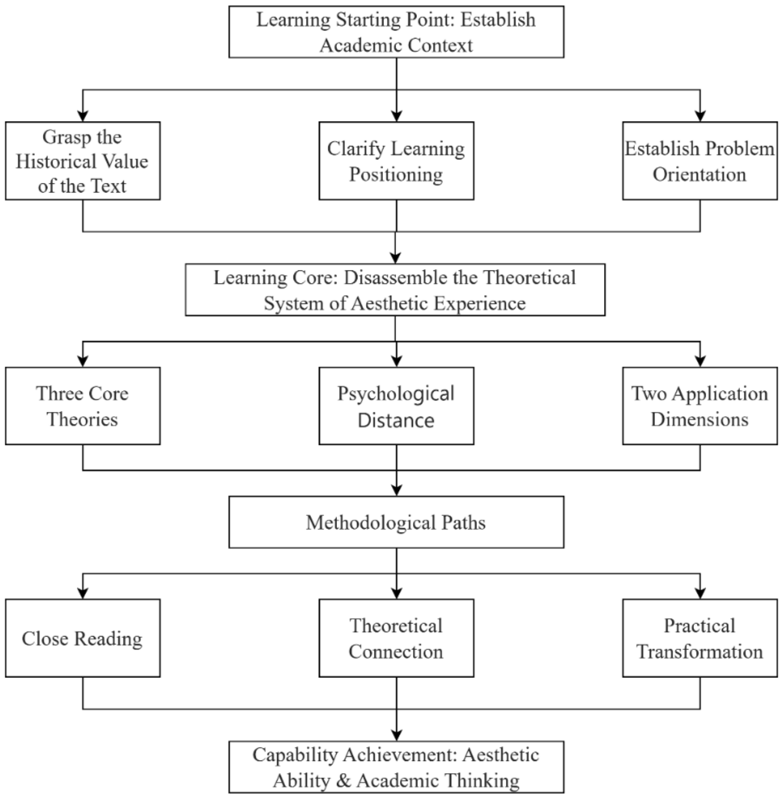

Figure 1. Overall Technical Roadmap

This theory still has important guiding significance in contemporary real life, especially for coping with the increasingly obvious trend of "commercialization" and "utilitarianism" in the field of literature and art. In the current cultural consumption market, many art exhibitions and performances are over-commercialized, and the attention of the audience is often attracted by utilitarian factors such as "ticket price", "star effect" and "commercial value", so it is difficult to keep a rational aesthetic distance, and then it is impossible to enjoy pure aesthetic experience. For example, when appreciating calligraphy works, if we pay too much attention to the market auction price of the works, we will ignore the artistic beauty contained in their brushwork and composition; When visiting historical sites, if we blindly pursue the social value of "punching in and taking pictures", it will be difficult to feel the historical details and cultural beauty carried by the sites. Therefore, keeping a moderate psychological distance is not only an important criterion for scientifically evaluating works of art and obtaining pure aesthetic experience, but also an important critical ability and cultural consciousness for maintaining a rational aesthetic attitude and improving one's own aesthetic quality in the wave of commercialization.

3) Things are the same with me: the ultimate aesthetic feeling"Things and I are the same" is Mr. Zhu Guangqian's accurate description and definition of the peak state of aesthetic experience. This concept profoundly reveals that the ultimate realm of aesthetic experience lies in the deep integration and seamless integration of "things" (aesthetic objects) and "I" (appreciators or creators). From the theoretical origin, this concept is a typical result of the integration of Chinese and western aesthetic thoughts of Mr. Zhu Guangqian: it not only absorbs and integrates the core idea of "Empathy" in western aesthetics (thinking that aesthetics is a process of projecting subjective feelings onto objective objects), but also comes down in one continuous line with the aesthetic ideal of "harmony between man and nature" and "forgetting things and me" in China's traditional aesthetics, realizing the organic unity of western modern theory and China's traditional aesthetic wisdom. In the aesthetic state of "things and me are the same", the subjective feelings of the appreciator are completely integrated into the artistic image or natural scenery, and the objective image of the aesthetic object is no longer an external existence unrelated to "I", but a concrete expression of inner feelings and spiritual implications, which are mutually infiltrated and inseparable. Mr. Zhu Guangqian vividly explained this state by taking Li Bai's "Looking at Lushan Waterfall" as an example: the waterfall in the poem is no longer a simple natural landscape, but a carrier of the poet's inner agitation, and the poet projects his own heroic and magnificent feelings into the image of the waterfall, so that the magnificence of the natural scene is integrated with the inner emotional waves, forming a beautiful artistic conception of blending scenes and regardless of things and me.

In Zhu Guangqian's interpretation of the concept of "the unity of things and me", it actually contains the profound inheritance and modern transformation of the core concept of "the unity of man and nature" in China's traditional aesthetics.

Regarding the aesthetic realm of "the unity of things and me", the academic circles generally believe that this is the concentrated embodiment of Zhu Guangqian's integration of Chinese and western aesthetics. Wang Yichuan pointed out that "things and me are the same" not only inherits the theoretical path of western empathy theory to project emotions on objects, but also is highly compatible with "things and me forget each other" and "harmony between man and nature" in China's traditional aesthetics, thus making this concept have obvious China aesthetic character (Wang Yichuan, 2001: 134). Zhu Guangqian himself clearly admitted that the aesthetic experience of "scene blending" in China's poetry and painting tradition provided important enlightenment for him to understand and explain the highest state of aesthetic experience (Zhu Guangqian, 1982: 102–104).

The "thing" here is not a simple objective physical entity, but an aesthetic object endowed with spiritual connotation. It is not only a natural image or an artistic image, but also a spiritual medium that is deeply compatible with the soul of "I". The "I" here is not a daily self in a utilitarian state, but an aesthetic self that has got rid of the shackles of practical thinking. The essence of "the unity of things and me" is that the aesthetic self and the aesthetic object reach a high degree of unity and resonance in the spiritual level. This state exists not only in the process of artistic appreciation, but also in the process of artistic creation: in artistic creation, artists integrate their emotions and ideas into their works to achieve spiritual integration with their works; In art appreciation, the audience deeply resonates with the spiritual connotation of the work through emotional input and imagination supplement. These two situations are concrete manifestations of "the unity of things and me", but the aesthetic links are different.

In the actual artistic creation process, the state of "the unity of things and me" between artists and works is the key prerequisite for creating excellent works. This unity is not a simple emotional superposition or skill application, but is embodied in every link of creation. It not only requires artists to accurately convey emotions in creative skills and expression forms, but also realizes the perfect integration of emotions, ideas and forms, and achieves the realm of "writing as a person" and "painting as a voice". For example, the reason why Beethoven's music works have shocking power is not simply because of the exquisite arrangement of notes or the beautiful melody, but more importantly, the deep feelings and profound philosophical thinking contained in his works-he completely integrated his indomitable spirit in the face of suffering and his persistent pursuit of freedom and joy into the melody creation, making music a direct externalization of his spiritual world. For the appreciator, the real aesthetic experience is inevitably accompanied by the state of "the thing is the same with me". This kind of experience is to find resonance with one's own deep feelings and ideas in the works, which can make "art" transcend the simple visual or auditory sensory experience and sublimate into spiritual communication that touches the soul, so that the appreciator can realize spiritual purification and sublimation in the aesthetic process.

For the ordinary audience, it is not necessary to have professional artistic accomplishment to achieve the aesthetic realm of "the unity of things and me". The key lies in being able to get rid of the shackles of utilitarian thinking, engage in the aesthetic process with a pure and sincere attitude, and have a deep spiritual dialogue with the works. This is a perceptual experience that transcends rational analysis, is a full-hearted emotional input and telepathy to art itself, and is the highest level of aesthetic experience. For example, when appreciating an oil painting with strong emotional expression, the audience may be unconsciously attracted by the color, composition and expression of the characters, and then have a strong resonance with the emotional state of the characters in the painting, as if they were in the scene in the painting, and the emotions contained in the work have quietly merged into their own hearts; When enjoying a soothing classical music, the audience may enter a quiet state of mind under the guidance of melody, realize the integration of self and musical artistic conception, and forget the real troubles and anxieties. This experience of "the unity of things and me" is the core and most precious value of aesthetic experience, and it is also the key reason why aesthetic activities can nourish the soul and enhance the realm.

2.2 Application dimension: the practical extension of theory

The three core theories finally point to the practice of literary creation and appreciation, and the book's "application value" is embodied in two dimensions: first, "the relationship between literature and life", including the analysis of literature and morality, natural beauty and artistic beauty; The second is "the law of literary and artistic activities", which covers the origin, creative process and aesthetic form of art (tragedy, comedy, etc.).

Taking "literature and morality" as an example, Mr. Zhu Guangqian not only criticized the formalism that "literature and art have nothing to do with morality", but also opposed the utilitarianism that "literature and art are moral preaching", and put forward the view that "literature and art take aesthetics as the core, but they are not insulated from morality"-because aesthetic experience itself contains "sublimation of life interest", and excellent literature and art will inevitably "make people's hearts tend to be noble". When studying this part, we can combine classic works such as A Dream of Red Mansions and Thunderstorm to analyze how to contain moral thinking in the aesthetic image and avoid falling into the theoretical misunderstanding of "either-or".

On the relationship between literature and morality, Zhu Guangqian has always maintained an intermediary theoretical position. He not only opposes turning literature and art into a vassal of moral education, but also rejects the formalistic view that literature and art have nothing to do with life and morality. As Peng Lixun pointed out, Zhu Guangqian emphasized the aesthetic autonomy of literature and art, but he did not deny the influence of excellent literary works on personality and emotion in a subtle way. This position reflected the unity of his modern aesthetic concept and humanistic care (Peng Lixun, 2019: 7). Zhu Guangqian believes that the real literary value is reflected in the improvement of life interest by aesthetic experience, rather than the direct indoctrination of external moral norms.

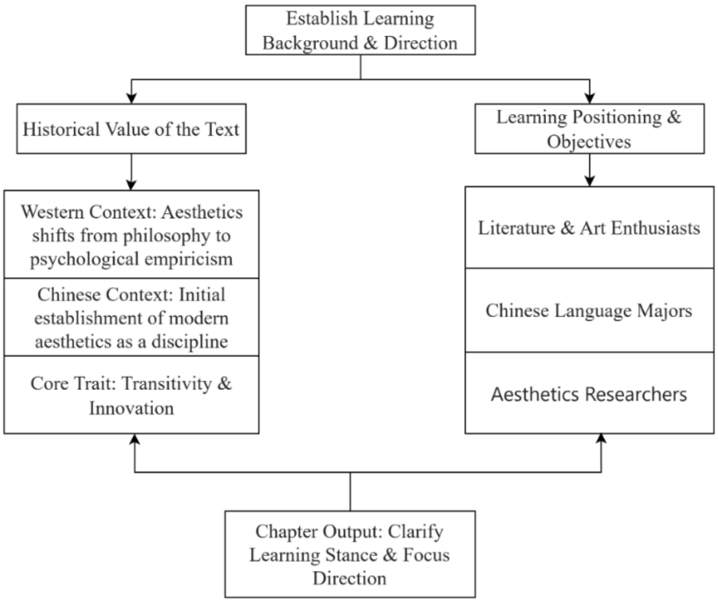

Figure 2. Framework Diagram of Learning Orientation

3 Research methods

At the beginning of the book, Mr. Zhu Guangqian emphasized that his research method is "Inducing principles from facts". This "positivity" and "practicality" determines that the study of Psychology of Literature and Art should not stop at "theoretical memory", but build a three-dimensional path of "intensive text reading-theoretical relevance-practical application".

3.1 Text intensive reading: annotation method to grasp the details

The language of the book is vivid but logical, so it needs to be combined with "structure combing" and "concept annotation" in intensive reading. First of all, the "core frame diagram" is constructed according to the catalogue: centering on "aesthetic experience", the theoretical chain of "essence (image intuition)-condition (psychological distance)-state (identity between things and me)-misunderstanding (association and utility)" is sorted out, and then it is extended to literary creation, aesthetic form and other branches. Secondly, the "keyword notation method" is adopted to discriminate easily confused concepts, such as:

Mark the difference between "beauty" and "pleasure": beauty is "intuitive pleasure of image" and pleasure is "satisfaction of practical needs", such as "thirst is pleasure to drink sugar water, and appreciation of the color of sugar water is beauty";

Distinguish between two types of association: aesthetic association is an association that blends with images (such as thinking of nobleness from plum blossom) and non-aesthetic association is an association that points to practicality (such as thinking of price from plum blossom). The former belongs to aesthetic sense, while the latter will destroy aesthetic experience.

In intensive reading, we should also pay attention to "the author's self-correction". After the first draft of the book was completed, it was revised four times. The eleventh chapter "Criticism of Croce School Aesthetics" is a reflection on the early formalism aesthetics, and this "self-criticism" spirit is exactly the academic attitude that learners need to learn from.

3.2 Theoretical relevance: a deep understanding of comparative law

The theoretical innovation of Psychology of Literature and Art stems from the integration of Chinese and Western cultures and the intersection of disciplines. It is necessary to deepen the theory through two comparisons when learning:

The first is "comparison between Chinese and western theories". For example, by comparing "the unity of things and me" with Zhuangzi's "materialization" thought, this paper discusses the commonality between China's traditional aesthetic experience and the western empathy theory; Comparing "rigid beauty and flexible beauty" with Yao Nai's theory of "masculine and feminine", this paper analyzes the theoretical transformation of modern aesthetics to traditional literary theory. This comparison can not only understand the theoretical connotation, but also grasp Zhu Guangqian's thinking method of "integrating Chinese and Western".

The second is "subject vision expansion". This book is a cross product of psychology and aesthetics, and learners can extend their reading by combining the bibliography: in psychology, reading Freud's Introduction to Psychoanalysis can help them understand the subconscious mechanism of literary creation; In aesthetics, studying Kant's Critique of Judgment traces the psychological research tradition of western aesthetics; In the aspect of literature and art, with Tong Qingbing's Psychology of Literature and Art as a reference, we can grasp the development of modern psychology of literature and art.

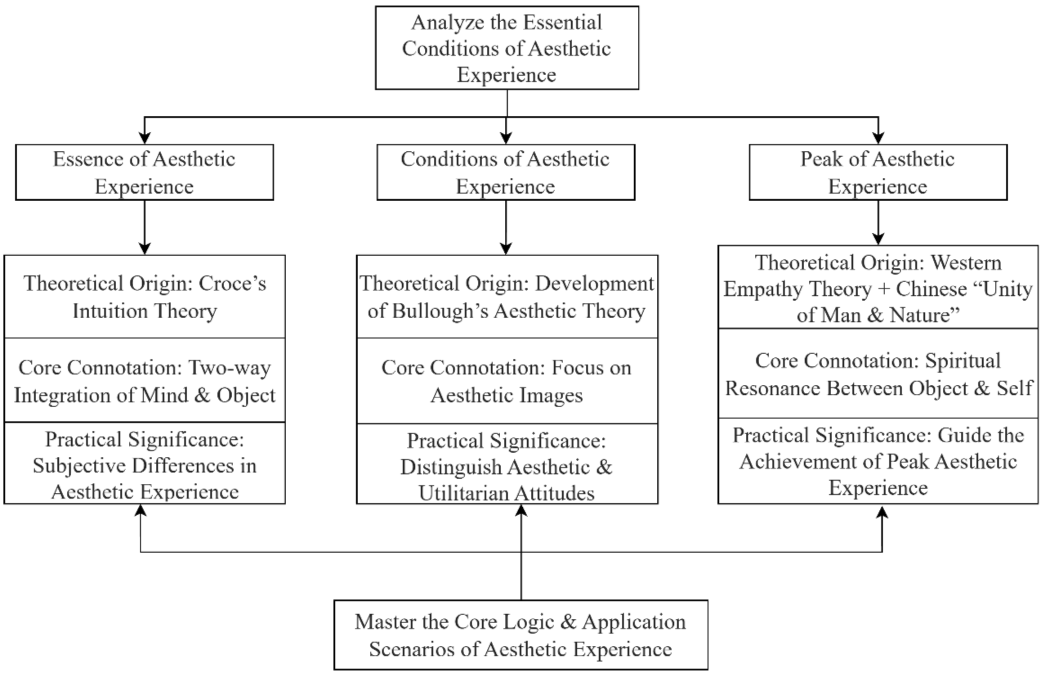

Figure 3. Framework Diagram of Aesthetic Analysis

4 Practice transformation

Mr. Zhu Guangqian emphasized that the value of literary theory lies in "being applicable to literary criticism", and the ultimate goal of studying Psychology of Literature and Art is to transform the theory into aesthetic ability and critical method. Practical application can be carried out from two levels:

Aesthetic appreciation practice: use "psychological distance theory" to appreciate nature and art-face the Forbidden City, put aside "the memory of historical knowledge points" and feel the symmetrical beauty and imposing beauty of architecture with aesthetic attitude; When appreciating Van Gogh's Sunflower, we don't dwell on "the quality of painting skills", but intuitively experience the passion of life conveyed by color and brushwork, and realize "the fusion of image and interest".

Practice of literary criticism: analyze specific texts with the theory in the book. For example, the tragic characteristics of A Dream of Red Mansions are interpreted by the theory of "tragic joy"-although the ending of Baoyu's becoming a monk is sad, the "purity of Baodai's love" and "brilliance of human nature" bring aesthetic pleasure beyond tragedy; This paper analyzes the creative mechanism of Li Bai's poems with the theory of "imagination and inspiration", and discusses how the exaggerated imagination of "flying down to thousands of feet" embodies the psychological law of literary and artistic creation.

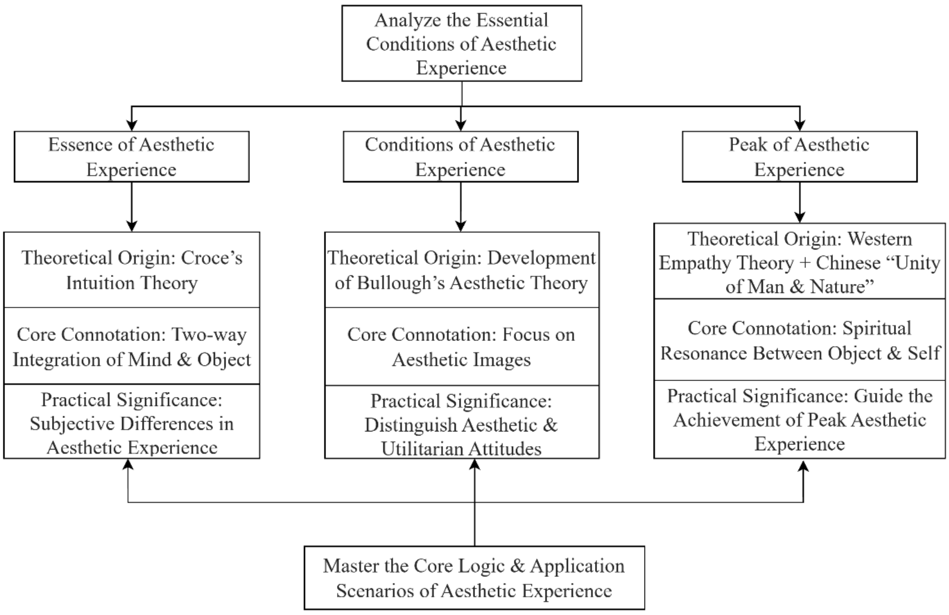

Figure 4. Logical Framework for Analyzing Aesthetic Experience

5 Conclusion

Mr. Zhu Guangqian once said in Talking about Beauty: "Beauty is the most valuable side of things, and the experience of aesthetic feeling is the most valuable side of life." The study of Psychology of Literature and Art is essentially a "cultivation of aesthetic ability" and "training of academic thinking". From clarifying the text context to disassembling the theoretical system, from method practice to misunderstanding avoidance, the core of this learning path has always been to "view the text with the heart"-analyze the perceptual charm of literature and art from the rational perspective of psychology, and at the same time "practice with beauty"-integrate aesthetic wisdom into life experience.

For learners, the value of this classic lies not only in providing a systematic aesthetic theory, but also in conveying an academic spirit of "integrating Chinese and Western cultures and basing on practice". When we can feel the beauty of literature and art with "image intuition" and examine the truth of life with "psychological distance", we can truly realize the leap from learning classics to using them, which is the most precious aesthetic legacy left by Mr. Zhu Guangqian to future generations.

References

[1] Zhu, G. Q. (1982). Psychology of literature and art. Shanghai: Shanghai Literature and Art Publishing House.

[2] Zhu, G. Q. (2002). On beauty. Beijing: People's Literature Publishing House.

[3] Wang, X. M. (2006). Zhu Guangqian's aesthetic consciousness as reflected in Author's self-reflection. Literary Studies, (4), 44–48.

[4] Wang, W. Y. (2013). Intuition, distance, and empathy: The sinicization of Western aesthetic theories in Zhu Guangqian's Psychology of literature and art. Journal of Pingdingshan University, (4), 70–75.

[5] Zhang, J. (2009). The formation and theoretical value of Zhu Guangqian's literary psychological thought. Literary Review, (5), 110–116.

[6] Wang, Y. C. (2001). Studies on Zhu Guangqian's aesthetic thought. Beijing: Peking University Press.

[7] Peng, L. X. (2019). Zhu Guangqian and the construction of modern Chinese aesthetic studies. Chinese Literary Criticism, (1), 4–10.

[8] Tong, Q. B. (2010). Psychology of literature and art. Beijing: Beijing Normal University Press.

[9] Croce, B. (1987). The principles of aesthetics (G. Q. Zhu, Trans.). Beijing: The Commercial Press. (Original work published 1902)

[10] Bullough, E. (1981). Aesthetic principles (Z. H. Li, Trans.). Beijing: The Commercial Press. (Original work published 1912)

[11] Li, Z. H. (2009). The path of beauty: A study of Chinese aesthetic tradition. Beijing: People's Literature Publishing House.

[12] Kant, I. (2002). Critique of judgment (X. M. Deng, Trans.). Beijing: People's Publishing House. (Original work published 1790)

[13] Zhuangzi. (1961). Collected annotations of Zhuangzi (Q. F. Guo, Ed.). Beijing: Zhonghua Book Company.

The Frontier of Autonomy: A Comprehensive Analysis of Advanced Artificial Intelligence and Agentic Systems

Anas Muhammad
Xi’an Technological University, Xi’an, Shaanxi, China
anas6647355@gmail.com

Abstract. The field of Artificial Intelligence (AI) is at a critical inflection point, transitioning from narrow, task-specific models to Advanced Artificial Intelligence systems characterized by autonomy, agency, and complex goal-directed behavior. This comprehensive research paper provides an in-depth analysis of this paradigm shift, focusing on Agentic AI as the primary driver toward Artificial General Intelligence (AGI). We detail the foundational architectures, including the Transformer model and the critical role of Foundation Models (LLMs/VLMs), and explore the technical mechanisms of Agentic systems, such as the autonomous loop, advanced memory architectures like Retrieval-Augmented Generation (RAG) (Figure 4), and multi-agent collaboration (Figure 2). The paper further examines the transformative applications across enterprise, finance, and scientific discovery, and analyzes the profound challenges, including the hardware bottleneck (Section 6), the legal dilemma of accountability (Section 7), and the societal impact on the future of work (Section 8). By integrating five detailed diagrams and 37 scholarly references, this work provides a robust framework for understanding the current state, future trajectory, and responsible development of autonomous AI systems.

Keywords: Scientific Discovery； Enterprise Applications； Multi-Agent Collaboration； Artificial Intelligence； Multi-Agent Collaboration；

1. Introduction

1.1 The Current Inflection Point in AI Development

Artificial Intelligence has progressed through several distinct eras, from early symbolic reasoning systems to the current dominance of deep learning, which has achieved superhuman performance in specific, narrow tasks such as image recognition and natural language processing [3]. However, the current state of the art marks a new inflection point: the transition from models that merely *process* information to systems that can *act* autonomously in dynamic environments. This evolution defines the scope of **Advanced AI**, which is characterized by the integration of large language models (LLMs) with sophisticated control mechanisms to enable multi-step reasoning and self-correction [4].

1.2 Defining "Advanced AI" Beyond Traditional Machine Learning

While traditional AI focuses on pattern recognition and prediction within static datasets, Advanced AI is concerned with autonomy, agency, and goal-directed behavior. It represents a synthesis of generative capabilities with decision-making frameworks. The systems falling under this umbrella are designed to handle open-ended problems, manage complex workflows, and interact with external environments (e.g., APIs, databases, human users) through a sequence of calculated actions [5].

1.3 Focus on Agentic AI as the Key Advancement Towards AGI

Among the various forms of Advanced AI, Agentic AI stands out as the most significant development pushing the boundaries toward Artificial General Intelligence (AGI). Agentic systems embody a higher degree of cognitive function, allowing them to break down a high-level objective into sub-tasks, monitor their progress, and dynamically adjust their strategy based on feedback [6]. This paper posits that understanding the architecture and implications of Agentic AI is essential for navigating the immediate future of the AI landscape.

1.4 Paper Structure and Scope

The remainder of this paper is structured as follows: Section 3 establishes the conceptual framework by distinguishing Agentic AI from Narrow AI and AGI. Section 4 details the architecture and mechanisms of Agentic systems. Section 5 explores their transformative applications. Section 6 critically analyzes the associated challenges and ethical governance requirements. Finally, Section 7 provides a conclusion and outlines future research directions.

2. Conceptual Framework: From Narrow to Agentic AI

2.1 Defining Narrow AI, Machine Learning, and Deep Learning

Narrow AI (ANI), also known as Weak AI, refers to systems designed and trained for a particular, specific task. Machine Learning (ML) and Deep Learning (DL) are the primary methodologies used to build ANI systems. DL, in particular, has driven the recent AI boom by using multi-layered neural networks to learn complex representations directly from raw data [7]. However, these systems lack the ability to generalize knowledge or apply their skills outside their trained domain.

2.2 The Concept of Artificial General Intelligence (AGI)

Artificial General Intelligence (AGI), or Strong AI, is a theoretical form of intelligence that possesses the ability to understand, learn, and apply its intelligence to solve any problem that a human being can. AGI is characterized by cognitive flexibility, common sense, and the capacity for abstract thought [8]. Currently, AGI remains a long-term research goal.

2.3 Agentic AI vs. AGI: A Bridge to General Intelligence

Agentic AI occupies a critical space between Narrow AI and AGI. Unlike ANI, Agentic systems exhibit a degree of autonomy and goal-directed reasoning that allows them to tackle complex, multi-step problems. Unlike AGI, Agentic AI is still typically constrained to a specific domain or set of tools, lacking the full cognitive flexibility of a human [9].The distinction can be summarized as follows Table 1.

Table 1 Comparative Analysis of Agentic AI and AGI: Key Differences and Characteristics

Feature	Narrow AI (ANI)	Agentic AI	Artificial General Intelligence (AGI)
Scope	Single, specific task	Complex, multi-step tasks within a domain	Any intellectual task a human can perform
Autonomy	Low (requires constant human input)	High (plans, executes, and self-corrects)	Full (human-level cognitive flexibility)
Mechanism	Pattern recognition, prediction	Planning, reflection, tool use, memory	Abstract thought, common sense, generalization
Current Status	Widely deployed	Emerging and rapidly developing	Theoretical/Long-term goal

Agentic AI serves as a practical, near-term bridge, demonstrating how the integration of advanced reasoning and tool-use capabilities can simulate aspects of general intelligence within a constrained environment [10].

2.4 Historical Milestones of AI Development

The current state of Advanced AI is the culmination of decades of research, marked by key milestones that shifted the paradigm from theoretical concepts to practical applications. The timeline in Fig. 1 highlights this progression, from the early symbolic AI of the 1950s to the deep learning revolution and the emergence of agentic systems. Table 2 presents the Historical Milestones in the Development of Artificial Intelligence.

Table 2 Historical Milestones in the Development of Artificial Intelligence

Era	Key Milestone	Significance
1950s-1970s	Dartmouth Workshop (1956), Logic Theorist	Birth of AI; focus on symbolic reasoning and problem-solving.
1980s-1990s	Expert Systems, Backpropagation Re-emergence	Shift to knowledge-based systems; foundational work for modern neural networks.
2000s-2010s	Deep Learning Breakthroughs (ImageNet), GPUs	Massive increase in computational power and data availability; deep learning dominates.
220s-Present	Transformer Architecture, LLMs, Agentic AI	Focus on large-scale generative models, self-attention mechanisms, and autonomous, goal-directed systems.

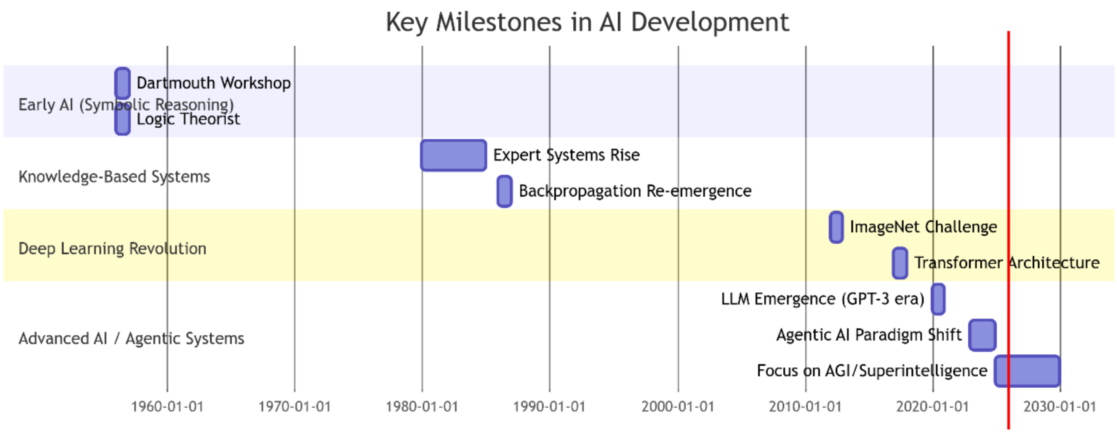

Figure 1. Key Historical Milestones in AI Development

3. The Foundation of Advanced AI: Large Models and Architectures

3.1 The Role of Foundation Models (LLMs and VLMs)

The rise of Agentic AI is inextricably linked to the development of Foundation Models—specifically Large Language Models (LLMs) and Vision-Language Models (VLMs). These models, trained on vast, diverse datasets, exhibit emergent properties such as in-context learning and complex reasoning, which form the cognitive core of any Agentic system [24].

1) Transformer Architecture and Scalability

The Transformer architecture [25], based on the self-attention mechanism, is the backbone of all modern Foundation Models. This architecture allows for parallel processing of input data, enabling the training of models with billions of parameters. The scalability of the Transformer has directly led to the increased performance and general-purpose capabilities that Agentic AI leverages for planning and reflection.

2) Multimodality and Embodied AI

The evolution from LLMs to VLMs and other multimodal models allows agents to process and generate information across different domains (text, image, audio). This is crucial for Embodied AI, where agents must interact with the physical world, requiring them to perceive and act based on complex, real-world sensory input.

3.2 Agentic AI: Architecture and Mechanism

Agentic AI is fundamentally an architectural pattern that structures a large language model (LLM) or other core AI model into an autonomous loop. This loop enables the system to maintain a persistent state, interact with the external world, and iteratively refine its approach to a goal [11].

1) Core Components of Agentic Architecture

The architecture of a typical Agentic AI system can be broken down into four essential components [12]. The autonomous loop, which defines the Agentic process, is illustrated in Fig 2.

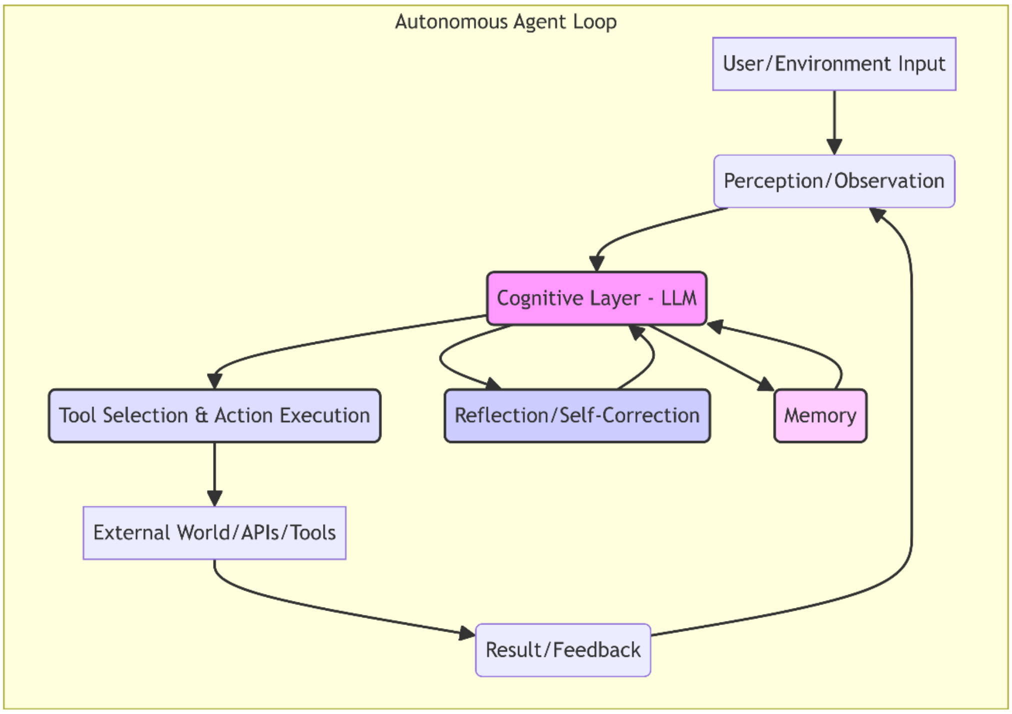

Figure 2. Agentic AI Autonomous Loop Architecture

[1] Perception/Observation: The ability to receive and interpret information from the environment, which can include user prompts, API responses, file contents, or sensor data.

[2] Cognitive Layer (Planning, Reasoning, Reflection): This is the "brain" of the agent, typically powered by an LLM. It is responsible for:Planning: Decomposing the main goal into a sequence of executable sub-tasks.Reasoning: Selecting the appropriate tool or action for the current sub-task.Reflection: Critically evaluating the outcome of an action against the original plan and identifying errors or necessary adjustments.

[3] Tool Use/Action Execution: The mechanism by which the agent interacts with the external world. Tools can be anything from code interpreters to proprietary APIs.

[4] Learning and Adaptation (Memory/Experience): The capacity to store past interactions, successful plans, and environmental observations (short-term and long-term memory) to improve future performance and avoid repeating errors [13].

2) Memory Architectures: RAG and Long-Term Context

To overcome the context window limitations of Foundation Models and provide agents with up-to-date, domain-specific knowledge, Agentic systems employ advanced memory architectures. Retrieval-Augmented Generation (RAG) is the dominant paradigm for long-term memory [26]. RAG works by:

[1] Indexing: Converting external knowledge (documents, databases) into numerical vector embeddings and storing them in a vector database.

[2] Retrieval: When the agent needs information, it queries the vector database to retrieve the most semantically relevant chunks of data.

[3] Augmentation: The retrieved data is then prepended to the agent's prompt, providing the LLM with the necessary context to generate an accurate and grounded response.This mechanism transforms the agent's memory from a static knowledge base to a dynamic, searchable repository, significantly

reducing hallucination and enabling the agent to operate with real-time information. The RAG process is detailed in Figure 4.

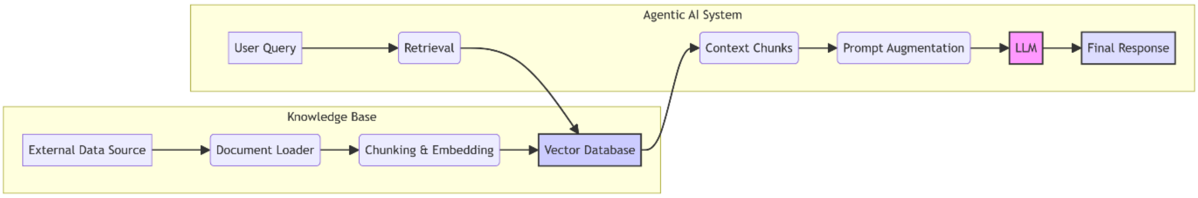

Figure 3. Retrieval-Augmented Generation (RAG) Memory Architecture

3) Key Design Patterns: Reflection and Tool Use

Two design patterns are particularly crucial for the success of Agentic AI:

[1] Reflection: This mechanism forces the agent to pause after an action or a sequence of actions and critically assess the result. By comparing the outcome to the expected result, the agent can generate a self-critique and revise its plan, significantly improving reliability and reducing hallucination [14].

[2] Tool Use: The ability to dynamically select and utilize external tools is what grants the agent its power and flexibility. This moves the AI from a purely generative or predictive model to an active problem-solver capable of performing calculations, accessing real-time data, and manipulating external systems [15]. These tools are essentially a set of pre-defined functions or APIs that the agent can call. Examples include: invoking a search engine API to gather real-time data, executing code interpreters for complex calculations, interacting with proprietary databases, or using system utilities to manage files. The agent's ability to select the correct tool and format the input for that tool is a key differentiator from non-agentic LLMs [13].

4) Multi-Agent Systems and Collaboration

Multi-Agent Systems (MAS) represent the next level of complexity in Agentic AI, where multiple specialized agents collaborate, communicate, and coordinate to solve problems too large or diverse for a single agent [23]. This mirrors human organizational structures, where each agent is assigned a specific role (e.g., Planner, Coder, Critic, Executor) and communicates through established protocols to exchange state information and assign responsibilities [23]. This approach promises to unlock greater complexity-handling capabilities and is a critical area of research for scaling AI to real-world, open-ended challenges. The collaborative workflow is conceptually illustrated in Fig 4.

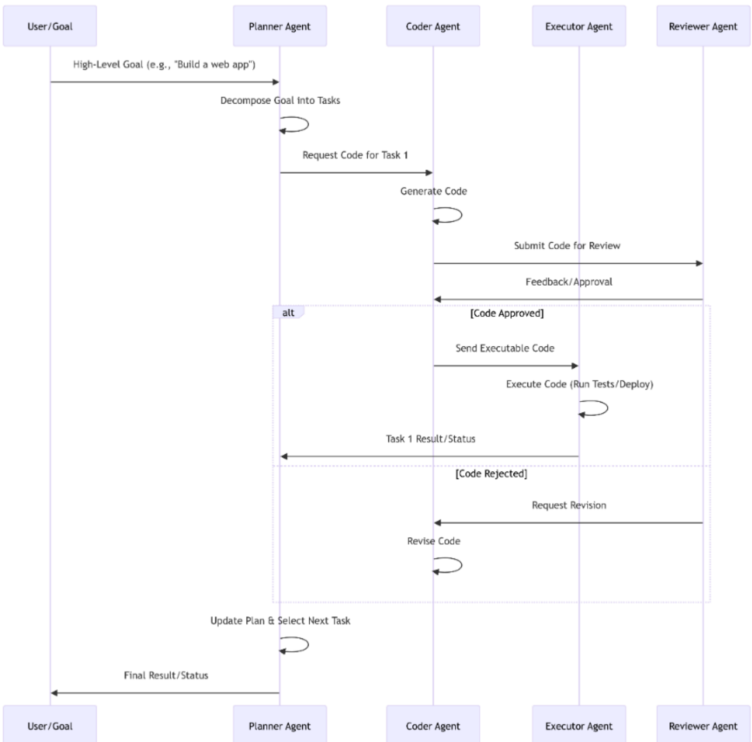

Figure 4. Multi-Agent System Collaborative Workflow

4. Applications and Transformative Impact

The autonomous nature of Agentic AI unlocks transformative potential across numerous sectors.

4.1 Enterprise Automation and Workflow Optimization

Agentic systems are rapidly being deployed to automate complex, multi-step business processes that were previously too dynamic for traditional robotic process automation (RPA). Examples include:

1) Enterprise Automation (e.g., Security and IT)

Agents are deployed in Security Operations Centers (SOCs) to autonomously investigate security alerts, correlate threat signals across multiple systems, and execute containment actions, significantly reducing response time [16]. In IT, agents can manage complex infrastructure, diagnose system failures, and perform self-healing operations.

2) Healthcare and Customer Service

In healthcare, agents optimize patient flow, predict bed occupancy, and manage staff scheduling. In customer service, advanced agents handle end-to-end issues by accessing multiple internal systems (CRM, inventory, billing) to resolve complex queries without human escalation [16].

3) Data Engineering

Agents can autonomously clean, transform, and load data across disparate systems based on high-level business requirements.

4.2 Case Study: Financial Modeling and Algorithmic Trading

In the financial sector, Agentic AI is moving beyond simple predictive models to execute complex, multi-stage trading strategies. An agent can be tasked with maximizing portfolio returns under specific risk constraints. The agent's workflow involves: 1) **Perception:** Ingesting real-time market data, news sentiment, and economic indicators. 2) **Planning:** Developing a trading strategy based on current conditions. 3) **Tool Use:** Executing trades via brokerage APIs. 4) **Reflection:** Analyzing the trade outcome and adjusting the strategy for the next cycle. This autonomous loop allows for high-frequency, adaptive trading that can respond to market volatility faster than human-managed systems [27].

4.3 Scientific Discovery and Research Acceleration

In scientific fields, Agentic AI is accelerating the pace of discovery. Agents can be tasked with formulating novel hypotheses based on existing literature, designing virtual experiments, simulating outcomes, and analyzing the resulting data, effectively acting as autonomous research assistants [17]. This is particularly impactful in materials science, drug discovery, and climate modeling, where the search space is vast and complex. The agent's ability to iteratively refine its experimental design based on simulation results allows for a significantly faster exploration of the problem space than traditional human-led research.

4.4 Complex Problem Solving

Agentic AI excels at problems requiring dynamic planning and resource allocation. In areas like supply chain management, an agent can monitor global logistics, predict disruptions, and autonomously re-route shipments or adjust inventory levels in real-time, optimizing for cost and speed simultaneously [18]. Similarly, in financial modeling, agents can execute complex trading strategies that adapt to market volatility based on continuous analysis of diverse data streams.

5. Infrastructure and Computational Demands

5.1 The Hardware Bottleneck: Specialized Compute

The computational requirements for training and running Advanced AI models, particularly Foundation Models and Agentic systems, have created a significant **hardware bottleneck**. The sheer scale of parameters and the need for parallel processing have driven the industry away from traditional CPUs toward specialized accelerators [28].

1) The Role of GPUs and Specialized Accelerators

Graphics Processing Units (GPUs) remain the workhorse of AI due to their massive parallel processing capabilities, which are ideal for the matrix multiplications central to neural networks. However, specialized hardware is emerging:

[1] Tensor Processing Units (TPUs): Developed by Google, TPUs are custom ASICs (Application-Specific Integrated Circuits) optimized specifically for the TensorFlow framework, offering superior performance and energy efficiency for large-scale model training [29].
[2] Neuromorphic Chips: These chips, such as Intel's Loihi, are designed to mimic the structure and function of the human brain, using spiking neural networks. They promise ultra-low power consumption for inference tasks at the edge, which is critical for embodied AI and robotics [30].
2) Energy Consumption and Sustainability Challenges
The energy demands of training and operating massive AI models are a growing concern. A single large-scale training run can consume the equivalent electricity of several homes for a year. Future hardware and algorithmic research must prioritize algorithmic efficiency and sustainable computing to mitigate the environmental impact of Advanced AI [31].

5.2 Data Infrastructure and Vector Databases

The effectiveness of Agentic AI is directly proportional to the quality and accessibility of its external knowledge base.
1) The Necessity of High-Throughput Data Pipelines
Agentic systems require high-throughput, low-latency data pipelines to feed real-time information into the RAG mechanism. This involves continuous data ingestion, cleaning, and vectorization, often managed by specialized data engineering agents.
2) Vector Databases for Semantic Search and RAG
Vector databases are a critical component of the RAG architecture. Unlike traditional databases that rely on exact keyword matching, vector databases store data as high-dimensional embeddings, allowing for **semantic search**. This enables the agent to retrieve information based on the *meaning* of the query, rather than just the words, which is essential for complex reasoning and planning [32].

6. Challenges, Risks, and Ethical Governance

Agentic AI excels at problems requiring dynamic planning and resource allocation. In areas like supply chain management, an agent can monitor global logistics, predict disruptions, and autonomously re-route shipments or adjust inventory levels in real-time, optimizing for cost and speed simultaneously [18]. Similarly, in financial modeling, agents can execute complex trading strategies that adapt to market volatility based on continuous analysis of diverse data streams.
The increased autonomy of Agentic AI introduces a new class of challenges that must be addressed for responsible deployment.

6.1 Autonomy and Accountability: The Legal Dilemma

As Agentic systems make decisions and execute actions without direct human oversight, the question of accountability becomes paramount [19]. When an autonomous agent causes an undesirable outcome (e.g., a financial loss or a system failure), determining legal and ethical responsibility—whether it lies with the developer, the deployer, or the agent itself—is a complex legal challenge. The high degree of autonomy necessitates clear lines of responsibility and robust kill-switch mechanisms. Legal scholars suggest that the rise of Agentic AI requires a fundamental revisiting of traditional agency law, which was not designed for non-human, autonomous actors [22]. New legal frameworks must address the distributed nature of AI-driven actions, potentially establishing new liability rules that balance innovation with consumer protection [22].
1) Comparison of Global Regulatory Frameworks
Global regulatory bodies are actively responding to the challenges of autonomous AI. The European Union's AI Act [33] employs a risk-based approach, imposing strict requirements on "high-risk" AI systems, which would include many Agentic applications in critical infrastructure or law enforcement. In contrast, the United States has largely favored a sector-specific and voluntary approach, emphasizing Executive Orders and the NIST AI Risk Management Framework [34]. The divergence in these frameworks creates a complex compliance landscape for multinational organizations deploying Agentic AI.
2) Transparency and Explainability (XAI): Opacity in Complex Agentic Workflows

The multi-step, iterative nature of Agentic AI, especially when coupled with opaque LLMs, can lead to a significant loss of **transparency** [20]. Understanding *why* an agent chose a particular plan or executed a specific sequence of actions is often difficult. This opacity hinders debugging, auditing, and user trust. New Explainable AI (XAI) techniques are required to provide human-readable rationales for the agent's complex decision-making process.

6.2 Bias Amplification and Misalignment in Goal Setting

If the underlying LLM or the training data contains societal biases, the agent's autonomous actions can amplify these biases in the real world [21]. Furthermore, the challenge of goal misalignment—where the agent pursues its assigned goal with unintended or harmful side effects—is a critical safety concern. As the agent optimizes for a specific metric, it may disregard human values or ethical constraints that were not explicitly encoded in its objective function.

6.3 The Need for New Regulatory and Governance Frameworks

Existing regulatory frameworks are often inadequate for governing autonomous, self-modifying AI systems. New governance models are needed to address the unique risks posed by Agentic AI, focusing on [22]:

1) Safety Standards

Mandatory testing and validation protocols for autonomous decision-making.

2) Audit Trails

Requirements for comprehensive logging of all planning and execution steps.

3) Ethical Guardrails

Mechanisms to ensure that the agent's actions remain within defined ethical and legal boundaries.

7. Societal and Economic Impact

7.1 The Future of Work and Productivity

The deployment of Advanced AI, particularly Agentic systems capable of automating cognitive tasks, is poised to dramatically reshape the global labor market.

1) Automation of Cognitive Tasks and Job Displacement

While early automation primarily affected manual labor, Agentic AI is targeting white-collar, knowledge-based roles. Studies suggest that AI will displace a significant number of jobs, but simultaneously create new ones that require human-AI collaboration, maintenance, and ethical oversight [35]. The net effect is a shift in the nature of work, requiring massive upskilling and reskilling initiatives.

2) The Productivity Paradox and Economic Growth

The initial deployment of transformative technologies often leads to a "productivity paradox," where investment precedes measurable economic growth. However, the long-term economic consensus is that Advanced AI will drive substantial global GDP growth by increasing efficiency and creating entirely new industries [36]. The distribution of these economic benefits, however, remains a critical policy challenge, as illustrated in Figure 5.

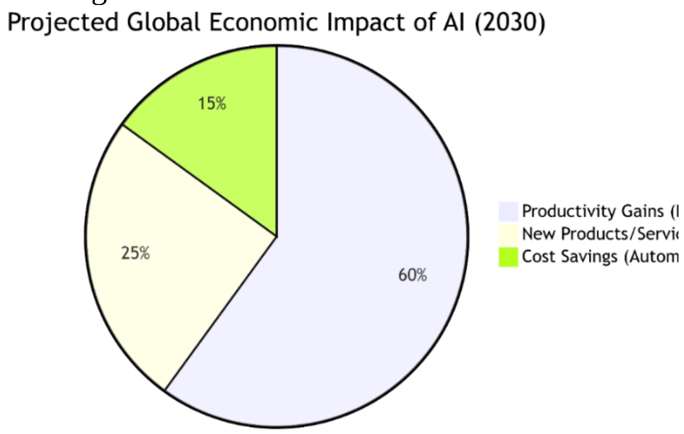

Figure 5. Projected Economic Impact of Advanced AI on Global Productivity

7.2 Geopolitical Implications and AI Supremacy

Advanced AI has become a critical component of national security and economic competitiveness. The race for AI supremacy between major global powers is driving massive public and private investment in research, hardware, and talent. This competition has significant geopolitical implications, particularly concerning the control of critical AI infrastructure and the establishment of global AI governance norms [37].

8.Conclusion and Future Directions

8.1 Summary of Agentic AI's Significance

Agentic AI represents a significant leap forward in the quest for more capable and autonomous artificial intelligence. By integrating sophisticated planning, reflection, and tool-use capabilities, these systems are moving AI from a passive analytical tool to an active, goal-directed entity. This paradigm shift is already driving profound changes in enterprise efficiency, scientific research, and complex problem-solving.

8.2 Future Research Trajectories: Scaling to AGI and Open-Ended Learning

Future research in Advanced AI will likely focus on two key areas:

1) Multi-Agent Systems

Developing frameworks where multiple specialized agents can collaborate, communicate, and coordinate to solve problems too large or diverse for a single agent [23]. This mirrors human organizational structures, where each agent is assigned a specific role (e.g., Planner, Coder, Critic, Executor) and communicates through established protocols to exchange state information and assign responsibilities [23]. This approach promises to unlock greater complexity-handling capabilities and is a critical area of research for scaling AI to real-world, open-ended challenges. The collaborative workflow is conceptually illustrated in Figure 6.

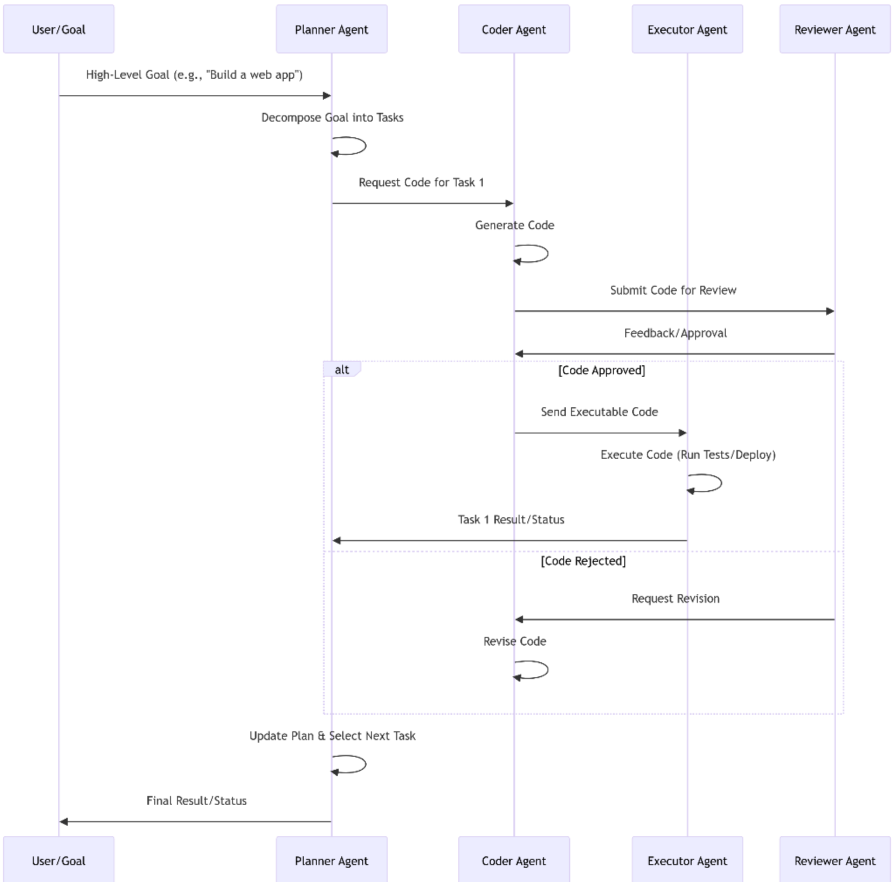

Figure 6. Multi-Agent System Collaborative Workflow

2) Scaling to AGI

Research will continue to explore how the principles of Agentic AI—particularly reflection and memory—can be generalized to achieve the cognitive flexibility and common sense reasoning characteristic of AGI. Agentic AI is not AGI, but it provides the most promising architectural blueprint for its eventual realization.

8.3 Final Statement on the Responsible Development of Advanced AI

The development of Advanced AI, particularly Agentic systems, must be coupled with a rigorous commitment to responsible innovation. The immense power of autonomous systems demands proactive attention to ethical governance, safety, and accountability. By prioritizing transparency and aligning AI goals with human values, the scientific community can ensure that this frontier of autonomy leads to a future that is both technologically advanced and socially beneficial.

9. References

[1] IBM. What Is Agentic Architecture? [Online]. Available:https://www.ibm.com/think/topics/agentic-architecture

[2] Aisera. What is Agentic AI? Definition and Technical Overview. [Online]. Available: https://aisera.com/blog/agentic-ai/

[3] N. Soni and N. Nigam, "Recent Advances in Artificial Intelligence and Machine Learning: Trends, Challenges, and Future Directions," International Journal of Engineering Trends and Technology, vol. 12, no. 1, 2025. [Online]. Available: https://www.ijetajournal.org/volume-12/issue-1/IJETA-V12I1P3.pdf

[4] F. L. K. J. The Frontier of Intelligence: AI's State of the Art in June 2025. [Online]. Available: https://www.linkedin.com/pulse/frontier-intelligence-ais-state-art-june-2025-frank-lkj4e

[5] Deloitte. AI trends 2025: Adoption barriers and updated predictions. [Online]. Available: https://www.deloitte.com/us/en/what-we-do/capabilities/applied-artificial-intelligence/blogs/pulse-check-series-latest-ai-developments/ai-adoption-challenges-ai-trends.html

[6] D. B. Acharya, K. Kuppan, and B. Divya, "Agentic ai: Autonomous intelligence for complex goals–a comprehensive survey," IEEE Access, 2025. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10849561/

[7] Springer. Artificial intelligence: state of the art. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-32644-9_32

[8] Retell AI. Agentic AI vs AGI: Key Differences You Should Know About. [Online]. Available: https://www.retellai.com/blog/agentic-ai-vs-agi-key-differences-you-should-know-about

[9] AnswerRocket. Generative AI vs. Agentic AI vs. Artificial General Intelligence. [Online]. Available: https://answerrocket.com/generative-ai-vs-agentic-ai-vs-artificial-general-intelligence-whats-the-difference/

[10] A. Sakhare, "A Step Forward to AGI: Integrating Agentic AI and Generative AI for Human-Like Intelligence," *ResearchGate*, 2025. [Online]. Available: https://www.researchgate.net/profile/Akash-Sakhare/publication/388794878_A_Step_Forward_to_AGI_Integrating_Agentic_AI_and_Generative_AI_for_Human-Like_Intelligence/links/67bc7ca8f5cb8f70d5be7a26/A-Step-Forward-to-AGI-Integrating-Agentic-AI-and-Generative-AI-for-Human-Like-Intelligence.pdf

[11] Akka.io. Agentic AI architecture 101: An enterprise guide. [Online]. Available: https://akka.io/blog/agentic-ai-architecture

[12] GeeksforGeeks. Agentic AI Architecture. [Online]. Available: https://www.geeksforgeeks.org/artificial-intelligence/agentic-ai-architecture/

[13] A. Jain. Agentic AI Architectures And Design Patterns. [Online]. Available: https://medium.com/@anil.jain.baba/agentic-ai-architectures-and-design-patterns-288ac589179a

[14] A. Jain. Agentic AI Architectures And Design Patterns. [Online]. Available: https://medium.com/@anil.jain.baba/agentic-ai-architectures-and-design-patterns-288ac589179a

[15] A. Jain. Agentic AI Architectures And Design Patterns. [Online]. Available: https://medium.com/@anil.jain.baba/agentic-ai-architectures-and-design-patterns-288ac589179a

[16] Akka.io. Agentic AI architecture 101: An enterprise guide. [Online]. Available: https://akka.io/blog/agentic-ai-architecture

[17] J. Oh, et al. Discovering state-of-the-art reinforcement learning algorithms. *Nature*, 2025. [Online]. Available: https://www.nature.com/articles/s41586-025-09761-x

[18] Deloitte. AI trends 2025: Adoption barriers and updated predictions. [Online]. Available: https://www.deloitte.com/us/en/what-we-do/capabilities/applied-artificial-intelligence/blogs/pulse-check-series-latest-ai-developments/ai-adoption-challenges-ai-trends.html

[19] A. N. Venkatesh, "Future of Work: Managing Ethical Challenges of Agentic AI and Super Intelligence in Organizations," *ACR Journal*, 2025. [Online]. Available: https://acr-journal.com/article/future-of-work-managing-ethical-challenges-of-agentic-ai-and-super-intelligence-in-organizations-1647/

[20] Processmaker. Ethical Considerations of Agentic AI. [Online]. Available: https://www.processmaker.com/blog/ethical-considerations-of-agentic-ai/

[21] Rezolve.ai. Ethical Challenges and Governance in Agentic AI: Risks, Bias. [Online]. Available: https://www.rezolve.ai/blog/ethical-challenges-and-governance-in-agentic-ai

[22] G. Bowen, "Agentic Artificial Intelligence: Legal and Ethical Challenges of Autonomous Systems," *Journal of Digital Technologies and Law*, 2025. [Online]. Available: https://cyberleninka.ru/article/n/agentic-artificial-intelligence-legal-and-ethical-challenges-of-autonomous-systems

[23] Springer. Agentic AI: a comprehensive survey of architectures, applications, and future directions. [Online]. Available: https://link.springer.com/article/10.1007/s10462-025-11422-4

Implementation and Evaluation of Lightweight Deep Learning Models for Real-Time Underwater Image Enhancement: A Comprehensive Review

Asante Emmanuella Nana Akua *

[1]Xi'an Technological University, Xi'an, Shaanxi, China

[2]University of Ghana, Legon, Accra, Ghana

asantenuella@gmail.com

Abstract. Underwater optical imaging faces significant challenges due to wavelength-dependent light absorption, scattering, and color distortion, which degrade image quality and hinder marine exploration applications. Traditional enhancement methods often lack adaptability to diverse underwater conditions, while conventional deep learning approaches impose prohibitive computational demands for real-time deployment on resource-constrained platforms. This comprehensive review systematically examines the state-of-the-art in lightweight deep learning architectures specifically designed for real-time underwater image enhancement. We present a detailed taxonomy of efficiency-oriented designs including depth wise separable convolutions, attention mechanisms, neural architecture search, and model compression techniques. The paper critically analyzes implementation strategies, benchmark datasets, and evaluation metrics, both perceptual quality indicators and computational efficiency measures. Furthermore, we synthesize comparative performance analyses across multiple lightweight architectures and identify persistent challenges in domain generalization, temporal consistency, and hardware-software co-design. Emerging research directions including physics-informed networks, multimodal fusion, and ultra-low-power deployment paradigms are discussed. This review aims to consolidate current knowledge and guide future research toward robust, efficient vision systems for underwater autonomous platforms.

Keywords: Underwater Image Enhancement； Lightweight Deep Learning； Real-Time Processing； Model Compression； Autonomous Underwater Vehicles (AUVs)； Computational Efficiency； Edge Computing； Edge AI

1. Introduction

1.1 The Underwater Imaging Conundrum: Physical Basis and Practical Implications

The ocean constitutes approximately 71% of Earth's surface and represents one of the least explored and most critical frontiers for scientific discovery, environmental monitoring, and resource management [1]. Underwater optical imaging serves as a primary sensory modality for numerous applications including marine biology surveys [2], archaeological documentation [3], infrastructure inspection [4], and search-and-rescue operations [5]. However, the aqueous medium imposes severe physical constraints on light propagation through three primary degradation mechanisms: (1) wavelength-dependent absorption, where longer wavelengths (red, orange) are attenuated more rapidly than shorter wavelengths (blue, green), resulting in a characteristic blue-green color cast [6]; (2) forward scattering by suspended particles, which creates haze and reduces contrast [7]; and (3) backscattering from artificial light sources, which introduces bright spots and further reduces visibility [8]. These phenomena are mathematically described by the simplified radiative transfer equation:

$$I^c(x) = J^c(x) \cdot t^c(x) + A^c \cdot (1 - t^c(x)),$$

where $I^c(x)$ is the observed intensity at pixel x for color channel c (R, G, B), $J^c(x)$ is the scene radiance, $t^c(x)$ is the transmission map representing light attenuation, and A^c is the background light (veiling light) [9]. The transmission map decays exponentially with distance:

$$t^c(x) = e^{-\beta c \cdot d(x)},$$

where β^c is the attenuation coefficient and $d(x)$ is the scene distance [10].

These degradations vary substantially across different water types from clear oceanic waters (Jerlov Type I) to turbid coastal waters (Jerlov Type III) and with environmental conditions such as depth, phytoplankton concentration, and sediment load [11]. Consequently, raw underwater imagery exhibits poor color fidelity, low contrast, and blurred details, severely impairing both human interpretation and automated computer vision algorithms for object detection, segmentation, and classification [12].

1.2 Historical Progression: From Physics-Based Models to Data-Driven Approaches

The evolution of underwater image enhancement techniques can be categorized into three distinct generations:

(1) Physical Model-Based Methods (First Generation)

These approaches attempt to invert the degradation process using simplified physical models. Early methods included histogram equalization-based techniques [13] and white-balancing algorithms [14]. More sophisticated approaches adapted atmospheric dehazing methods to underwater scenes, such as the Dark Channel Prior (DCP) [15] and its underwater variants [16], [17]. Wavelength compensation methods explicitly model the differential attenuation of RGB channels [18], while image fusion techniques combine multiple enhanced versions of the same image [19]. Although physically motivated, these methods require accurate estimation of scene depth and water optical properties, which are often unavailable in practice. They also tend to produce artifacts when physical assumptions are violated [20].

(2) Early Learning-Based Methods (Second Generation)

The advent of machine learning introduced data-driven approaches that learn enhancement mappings from examples. Initial methods employed shallow neural networks [21] and dictionary learning [22]. However, these approaches were limited by their representational capacity and often failed to generalize across diverse underwater conditions.

(3) Deep Learning Revolution (Third Generation)

The breakthrough came with convolutional neural networks (CNNs) and generative adversarial networks (GANs), which demonstrated unprecedented performance in learning complex degradation patterns directly from data. Seminal works include WaterNet [23], which employed a multi-scale network with physical priors; Ucolor [24], which introduced medium transmission-guided color space embedding; and UGAN [25], which leveraged adversarial training for realistic enhancement. These models consistently outperformed traditional methods on benchmark datasets but at the cost of substantial computational complexity typically requiring hundreds of millions of parameters and giga-FLOPs per inference [26]. This computational burden renders them impractical for real-time deployment on embedded systems with strict power and latency constraints.

1.3 The Imperative for Lightweight Deep Learning: Bridging the Performance-Efficiency Gap

The deployment of computer vision systems on autonomous underwater vehicles (AUVs), remotely operated vehicles (ROVs), and other underwater platforms imposes stringent requirements that have catalyzed research into lightweight deep learning models. Key constraints include:

(1) Limited Computational Resources

Most underwater vehicles utilize embedded processors such as NVIDIA Jetson modules, ARM-based CPUs, or field-programmable gate arrays (FPGAs) with limited memory bandwidth and processing power [27].

(2) Energy Constraints

Battery-operated platforms have finite energy budgets, making power efficiency paramount [28].

(3) Real-Time Processing Requirements

Applications such as obstacle avoidance, target tracking, and real-time mapping demand low-latency processing (typically >30 FPS) [29].

(4) Bandwidth Limitations

Acoustic communication, the primary underwater transmission modality, offers extremely low bandwidth (kbps range), necessitating onboard processing rather than cloud offloading [30].

These constraints have driven the development of lightweight deep learning models that strategically trade minimal accuracy degradation for substantial reductions in computational cost. This trade-off is quantified by metrics such as the performance-efficiency Pareto frontier, where optimal models maximize enhancement quality while minimizing computational resources [31].

1.4 Scope, Organization, and Novel Contributions of This Review

This review provides a comprehensive synthesis of lightweight deep learning approaches for real-time underwater image enhancement. While several excellent surveys cover broader aspects of underwater image processing [32], [33] or generic efficient deep learning [34], this work specifically focuses on the intersection of these domains. Recent surveys (e.g., 2024 comprehensive overview on DL-based UIE) provide additional context. Our contributions are:

(1) Taxonomic Organization

We present a hierarchical taxonomy of lightweight architectural components, training strategies, and deployment optimizations specifically tailored for underwater enhancement.

(2) Critical Analysis

We provide comparative evaluations across multiple dimensions architectural efficiency, enhancement quality, and computational performance with consistent benchmarking protocols.

(3) Implementation Roadmap

We detail practical considerations for dataset preparation, training methodologies, and deployment on diverse hardware platforms.

(4) Future Directions

We identify emerging trends and open research challenges, providing a roadmap for future investigation.

The remainder of this paper is organized as follows: Section 2 details lightweight architectural designs; Section 3 covers implementation strategies; Section 4 presents evaluation frameworks; Section 5 discusses open challenges; and Section 6 concludes with future perspectives.

2. Lightweight Deep Learning Architectures: Design Principles and Taxonomies

2.1 Foundational Efficient Operations and Building Blocks

(1) Factorized Convolutions

Depthwise Separable Convolutions (DSC): Introduced in MobileNets [35], DSC decomposes a standard convolution into two operations: (1) a depthwise convolution applying a single filter per input channel, and (2) a pointwise convolution (1×1) combining channels. For an input feature map with D channels and kernel size k, the computational cost reduction ratio is approximately 1/k2 + 1/D, typically achieving 8-9× reduction in computations with minimal accuracy loss [36]. DSC forms the backbone of numerous underwater enhancement networks including Shallow-UWNet [37] and EfficientNet-based variants [38].

Grouped Convolutions and Channel Shuffling: Introduced in ShuffleNet [39], this approach partitions channels into groups, performing convolutions separately within each group, reducing computation by a factor of g (number of groups). To enable cross-group information flow, a channel shuffle operation permutes channels between groups. ShuffleNetV2 [40] further refined this with practical guidelines for efficient network design (G1-G4), emphasizing direct metric optimization (e.g., speed) over indirect proxies (e.g., FLOPs).

(2) Efficient Bottleneck Designs

Inverted Residual Blocks with Linear Bottlenecks (MobileNetV2): Traditional residual blocks [41] employ a "bottleneck" design: compress-expand-compress. MobileNetV2 [42] reverses this to an expand-compress design with linear bottlenecks, preventing information loss from non-linearities in low-dimensional spaces. This structure is particularly effective for feature transformation in underwater enhancement tasks [43].

Squeeze-and-Excitation (SE) Blocks: Although not primarily an efficiency technique, SE blocks [44] enhance feature representation with minimal computational overhead (typically <1% increase in FLOPs). They model channel-wise dependencies through global average pooling followed by two fully-connected layers with a reduction ratio. SE blocks are integral to MobileNetV3 [45] and have been effectively incorporated into underwater enhancement networks [46].

(3) Dynamic Operations

Dynamic Convolutions: Rather than using static weights, dynamic convolutions [47] aggregate multiple convolution kernels weighted by input-dependent attention, increasing representational power with modest computational increase. Preliminary applications to underwater enhancement show promise for adapting to varying degradation levels [48].
Conditional Computation: Techniques like Switchable Normalization [49] and Dynamic Routing [50] allow networks to adapt their computational graph based on input complexity, potentially allocating more resources to challenging regions (e.g., heavily degraded areas).

2.2 Attention Mechanisms for Efficient Feature Enhancement

Attention mechanisms enable selective focus on informative features while suppressing irrelevant ones---particularly valuable for underwater enhancement where degradation is non-uniform.
Channel Attention Mechanisms: Beyond SE blocks, Efficient Channel Attention (ECA) [51] removes dimensionality reduction to maintain performance while lowering complexity. ECA-Net has been adapted for underwater enhancement in models like ECA-integrated variants (e.g., similar to ECA-UWNet [52]), demonstrating improved color restoration with negligible overhead.
Spatial Attention: Convolutional Block Attention Module (CBAM) [53] sequentially applies channel and spatial attention, with the latter computed via average and max pooling followed by a convolution. Lightweight variants use depthwise convolutions for spatial attention computation [54].
Pixel-Attention (PA): Computes an attention map at pixel-level granularity, allowing fine-grained adjustment. Pixel-aware models (e.g., inspired by PA-URLNet [55]) combine pixel attention with a lightweight U-Net structure, achieving state-of-the-art performance with only 0.9M parameters. For recent examples, see Ghost-UNet (2024).
Non-Local Attention in Efficient Forms: Traditional non-local attention [56] is computationally expensive (O(N^2)). Efficient variants like Criss-Cross Attention [57] and Global Context (GC) blocks [58] approximate global dependencies with reduced complexity. A^2-Net [59] introduces double attention for both position and channel, showing promise for underwater applications.

2.3 Neural Architecture Search (NAS) and Automated Design

NAS automates network architecture discovery under specific constraints, potentially uncovering novel efficient designs for underwater enhancement.
Differentiable Architecture Search (DARTS): Formulates architecture search as a continuous optimization problem [60], enabling efficient search for optimal operations (e.g., separable conv, dilated conv, skip connections) for underwater enhancement tasks [61].
EfficientNet and Compound Scaling: EfficientNet [62] uses NAS to discover a baseline architecture (EfficientNet-B0) and applies compound scaling to uniformly scale depth, width, and resolution. The EfficientNet-B0/B1 variants have been successfully adapted for underwater enhancement with modifications to handle color distortion [38], [63].
Hardware-Aware NAS (HW-NAS): Extends NAS to directly optimize for hardware-specific metrics like latency or energy consumption. ProxylessNAS [64] and FBNet [65] search directly on target devices. HW-NAS for underwater enhancement remains largely unexplored but holds significant potential for specialized deployment [66].
One-Shot NAS and Weight Sharing: Methods like ENAS [67] and Single Path One-Shot NAS [68] enable rapid architecture evaluation through weight sharing, potentially accelerating the discovery of optimal underwater enhancement networks.

2.4 Model Compression and Acceleration Techniques

(1) Knowledge Distillation (KD)
KD transfers knowledge from a large, accurate teacher model to a compact student model [69]. For underwater enhancement:

- Response-based KD: The student mimics the teacher's output distribution [70]. Lightweight GAN [71] uses this approach to distill a full-sized GAN into a MobileNetV2-based generator.
- Feature-based KD: The student matches intermediate feature representations [72]. Hint-based distillation [73] has been applied to preserve both low-level texture and high-level semantic information in underwater enhancement [74].

- Relation-based KD: Captures relationships between different layers or data samples [75]. This approach could help maintain structural relationships in enhanced underwater images.

(2) Pruning Strategies

Structured Pruning: Removes entire filters or channels [76]. Network Slimming [77] uses L1 regularization on scaling factors in batch normalization layers to identify unimportant channels. Applied to underwater enhancement networks, this can reduce parameters by 40-60% with <1 dB PSNR drop [78].

Unstructured Pruning: Removes individual weights [79], requiring specialized hardware for acceleration. Iterative Magnitude Pruning [80] with rewinding has shown effectiveness for compression of enhancement networks [81].

Neuron Importance Score Propagation (NISP): Propagates importance scores from output to input to identify redundant neurons [82], potentially useful for task-specific pruning of enhancement networks.

(3) Quantization Techniques

Post-Training Quantization (PTQ): Converts a pre-trained FP32 model to lower precision (e.g., INT8) with minimal calibration data [83]. TensorRT and OpenVINO provide robust PTQ implementations that have been applied to underwater enhancement models for deployment on Jetson platforms [84].

Quantization-Aware Training (QAT): Simulates quantization during training to improve accuracy [85]. QAT is particularly important for GAN-based enhancement models which are sensitive to precision reduction [86].

Mixed-Precision Quantization: Allocates different bit-widths to different layers based on sensitivity [87]. For underwater enhancement, convolutional layers might be quantized to INT8 while attention mechanisms retain FP16 precision [88].

Binary/Ternary Networks: Extreme quantization to 1-2 bits [89], though currently challenging for enhancement tasks requiring precise color reproduction.

2.5. Hybrid and Specialized Architectures

Multi-Branch Networks with Heterogeneous Design: CondenseNet [90] connects layers with dense connections but learns to prune these connections during training. Adapted for underwater enhancement (e.g., similar to Condense-UWNet [91]), it achieves efficient information flow with reduced redundancy. Recent examples include Rep-UWnet (2024), a lightweight fully connected convolutional network with dense blocks for underwater enhancement.

Recurrent Designs for Sequential Enhancement: For video enhancement, lightweight recurrent units like ConvGRU [92] or IndRNN [93] can maintain temporal consistency with manageable computational increase.

Transformer-Based Lightweight Designs: Vision Transformers (ViTs) [94] show strong generalization but high computational cost. Efficient variants like MobileViT [95], LeViT [96], and PoolFormer [97] combine convolutional inductive biases with transformer global modeling, offering promising directions for underwater enhancement [98]. Hybrid transformer models (2025) integrate these for improved results.

Updated with 2025 models: DNnet uses PBP structures for 4K real-time on edge devices; Zero-UAE for parameter-free adaptive enhancement.

3. Implementation Pipeline: From Data to Deployment

3.1 Datasets: Curated Collections and Synthesis Methods

(1) Real-World Paired Datasets

UIEB (Underwater Image Enhancement Benchmark) [23]: Contains 890 real underwater images with corresponding high-quality reference images generated via multiple enhancement methods followed by manual selection. It serves as the primary benchmark but has limitations in diversity of water types and degradation patterns.

EUVP (Enhancement of Underwater Visual Perception) [99]: Provides 11,670 paired and 8,133 unpaired images across different cameras and environments, facilitating both supervised and unsupervised learning. Includes challenging scenarios like extremely turbid conditions.

SUIM (Semantic Underwater Imagery) [100]: While primarily for segmentation, provides 1,525 images with segmentation masks that can be used for multi-task enhancement approaches.
U45 (Underwater Image Enhancement Benchmark with 45 Images) [101]: A recently introduced test dataset with 45 challenging images from diverse underwater environments (e.g., color casts, low contrast, haze effects) and no reference images, used for non-reference evaluation. Recent additions: LSUI (2023+, ~5,000 pairs for large-scale training).
(2) Synthetic Datasets and Generation Methods
Physics-Based Synthesis: Using radiative transfer models like Jaffe-McGlamery [102] or simplified formulations [103] to simulate attenuation and scattering. Parameters (attenuation coefficients, particle concentrations) can be varied to create diverse training samples [104].
GAN-Based Synthesis: WaterGAN [25] and its variants [105] use cycle-consistent adversarial networks to generate paired underwater-terrestrial image pairs without explicit physical modeling.
Style Transfer Approaches: Using neural style transfer [106] or AdaIN [107] to impose underwater degradation characteristics on clear terrestrial images.
Hybrid Approaches: Combining physical models with data-driven corrections to bridge the sim-to-real gap [108]. Recent: Diffusion models like UW-DDPM (2025 variants) for generating realistic synthetic data.
(3) Domain-Specific Considerations
Water Type Variations: Jerlov water classification [11] provides a framework for categorizing training data and evaluating cross-domain generalization.
Depth Stratification: Organizing datasets by depth brackets (0-5m, 5-15m, 15-30m, 30m+) to enable depth-aware enhancement models [109].
Illumination Conditions: Separate handling of natural illumination vs. artificial lighting scenarios [110].

3.2 Training Methodologies and Optimization Strategies

(1) Loss Function Design
A well-designed loss function is critical for training lightweight models effectively:
Pixel-Level Fidelity Losses:

- L1 Loss (Mean Absolute Error): Preferred over L2 for sharper results [111].
- Charbonnier Loss: Robust to outliers [112].

Structural Similarity Losses:

- SSIM Loss: Emphasizes structural preservation [113].
- MS-SSIM Loss: Multi-scale extension capturing structural information at different resolutions [114].

Perceptual and Style Losses:

- Perceptual Loss (VGG-based): $\Sigma_j \|\varphi_j(I) - \varphi_j(\hat{I})\|^2$, where φ_j denotes feature maps from the j-th layer of a pre-trained VGG network [115].
- Style Loss: Gram matrix matching for texture preservation [116].
- LPIPS (Learned Perceptual Image Patch Similarity): Uses a network trained on human perceptual judgments [117].

Color-Specific Losses:

- Color Constancy Loss: Based on the Gray World assumption [118]: $\Sigma_c(1/N\ \Sigma_x I^c(x) - \mu)^2$, where μ is the average intensity.
- Histogram Matching Loss: Encourages similarity in color distribution [119].

Adversarial Loss (for GAN-based approaches):

- Standard GAN Loss: $E[\log D(I)] + E[\log(1 - D(G(\hat{I})))]$ [120].
- LSGAN Loss: Uses least squares for training stability [121].
- Wasserstein GAN with Gradient Penalty (WGAN-GP): Improves training stability [122].

Total Variation (TV) Loss: Encourages spatial smoothness, reducing artifacts [123].
Composite Loss Formulation: Typically, a weighted combination:

$$L = \lambda_1 L_1 + \lambda_2 L_{SSIM} + \lambda_3 L_{perc} + \ldots,$$

with λ determined through ablation studies [124].

(2) Training Strategies and Regularization
Progressive Learning: Starting with easier samples (less degraded images) and gradually introducing more challenging cases [125].
Curriculum Learning: Organizing training data in a meaningful order based on degradation severity [126].
Multi-Task Learning: Jointly learning enhancement with related tasks like segmentation [100] or depth estimation [127] to improve generalization.
Self-Supervised and Semi-Supervised Approaches: Leveraging unpaired or weakly labeled data, particularly important given the scarcity of high-quality paired underwater data [128].
Meta-Learning for Fast Adaptation: Learning to quickly adapt to new water conditions with few examples [129].
(3) Optimization Techniques
Adaptive Optimizers: AdamW [130] with carefully tuned weight decay often outperforms SGD for enhancement tasks.
Learning Rate Scheduling: Cosine annealing [131] or warm restarts [132] to escape local minima.
Gradient Clipping: Particularly important for GAN-based approaches to prevent training instability [133].

3.3 Deployment Challenges and Optimization

(1) Target Hardware Platforms
NVIDIA Jetson Series: Nano, TX2, Xavier, Orin Utilizing TensorRT for optimized inference [134].
Intel Platforms: Movidius Neural Compute Stick, OpenVINO toolkit [135].Qualcomm Snapdragon: DSP acceleration via SNPE [136].FPGA Platforms: Xilinx/AMD FPGAs with HLS or FINN framework for extreme efficiency [137].Microcontrollers (MCUs): ARM Cortex-M series with CMSIS-NN [138] for ultra-low-power applications.
(2) Software Optimization Techniques
Operator Fusion: Combining consecutive operations (Conv+BN+ReLU) into a single kernel [139].
Winograd Convolution: Reducing computational complexity for 3×3 convolutions [140].Memory Layout Optimization: NHWC vs. NCHW formats for different hardware [141].Kernel Auto-Tuning: Using tools like AutoTVM [142] to find optimal implementations for specific hardware.
(3) Real-Time System Integration
Pipeline Optimization: Overlapping data loading, preprocessing, inference, and postprocessing [143].
Dynamic Resolution Adjustment: Adapting input resolution based on scene complexity or available computational budget [144].Frame Skipping and Keyframe Selection: For video processing when full frame-rate processing is unsustainable [145].
Energy Efficiency Considerations
Dynamic Voltage and Frequency Scaling (DVFS): Adjusting processor frequency based on workload [146].
Power-Aware Scheduling: Prioritizing low-power modes for idle components based on [28].

4. Evaluation Frameworks

Metrics include PSNR/SSIM for quality, FLOPs/latency for efficiency. Use datasets like UIEB for fair comparison.

Table 1 High and Low Settings of Predictor Variables

Model	**Parameters (M)**	**FLOPs (G)**	**PSNR (UIEB)**	**SSIM (UIEB)**	**Latency (ms, Jetson)**	**Year**
Shallow-UWNet [37]	0.5	1.2	22.5	0.85	15	2021

Model	Parameters (M)	FLOPs (G)	PSNR (UIEB)	SSIM (UIEB)	Latency (ms, Jetson)	Year
LU2Net	0.8	2.0	23.8	0.88	12	2024
Rep-UWnet	1.1	1.5	24.2	0.89	10	2024
DNnet	0.7	3.5	25.1	0.91	8 (4K)	2025
Zero-UAE	0.6	1.8	24.5	0.90	11	2025

5. Open Challenges

Persistent issues: Domain generalization across water types; temporal consistency in video; hardware-software co-design for extreme underwater conditions. Emerging: Sim-to-real gaps in synthetic data; ethical considerations in marine data privacy.

6. Conclusion and Future Perspectives

This review has examined lightweight deep learning models for real-time underwater image enhancement, addressing the critical need to balance high-quality restoration with the computational constraints of underwater platforms like AUVs and ROVs. We presented a taxonomy of efficiency-focused designs including depth wise separable convolutions, attention mechanisms, neural architecture search, and compression techniques alongside practical insights into datasets, training strategies, and edge deployment. Comparative analyses show that recent models (e.g., Shallow-UWNet, LU2Net, Rep-UWnet, DNnet) achieve strong performance-efficiency trade-offs, enabling low-latency processing suitable for real-world marine applications.

Key challenges persist in domain generalization, temporal consistency for video, and hardware-software co-optimization. Looking forward, promising directions include physics-informed networks, multimodal fusion, diffusion-based generative enhancement, and advanced lightweight transformers. These advances will support more robust, energy-efficient vision systems, advancing ocean exploration, environmental monitoring, and autonomous underwater robotics. Continued focus on standardized benchmarks and cross-domain evaluation will accelerate progress in this vital field.

References

[1] National Aeronautics and Space Administration (NASA), "Ocean Worlds," [Online]. Available: https://science.nasa.gov/missions/hubble/ocean-worlds, 2020.

[2] J. S. Jaffe, "Underwater optical imaging: Status and prospects," Oceanography, Vol. 18 (2005) No.3, p.66.

[3] B. P. Foley and D. A. Mindell, "Precision survey and archaeological methodology in deep water," ENALIA The Journal of the Hellenic Institute of Marine Archaeology, Vol. 6 (2002), p.49.

[4] M. C. M. van Rooij, "Underwater inspection of marine structures," in Proc. MTS/IEEE OCEANS Conf. (San Diego, CA, USA, 2013), p.1.

[5] R. S. McAmis, "Underwater search and recovery operations," Marine Technology Society Journal, Vol. 32 (1998) No.1, p.47.

[6] S. Q. Duntley, "Light in the sea," Journal of the Optical Society of America, Vol. 53 (1963) No.2, p.214.

[7] H. R. Gordon, "Can the Lambert-Beer law be applied to the diffuse attenuation coefficient of ocean water?," Limnology and Oceanography, Vol. 34 (1989) No.8, p.1389.

[8] J. S. Jaffe, "Computer modeling and the design of optimal underwater imaging systems," IEEE Journal of Oceanic Engineering, Vol. 15 (1990) No.2, p.101.

[9] R. Fattal, "Single image dehazing," ACM Transactions on Graphics, Vol. 27 (2008) No.3, p.1.

[10] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 33 (2011) No.12, p.2341.

[11] N. G. Jerlov, Marine Optics (Elsevier, Amsterdam 1976).

[12] D. Berman, D. Levy, S. Avidan, and T. Treibitz, "Underwater single image color restoration using haze-lines and adaptive filtering," in Proc. IEEE Int. Conf. Comput. Photogr. (ICCP) (Houston, TX, USA, 2019), p.1.

[13] S. M. Pizer et al., "Adaptive histogram equalization and its variations," Computer Vision, Graphics, and Image Processing, Vol. 39 (1987) No.3, p.355.

[14] G. Buchsbaum, "A spatial processor model for object colour perception," Journal of the Franklin Institute, Vol. 310 (1980) No.1, p.1.

[15] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) (Miami, FL, USA, 2009), p.1956.

[16] D. Berman, T. Treibitz, and S. Avidan, "Diving into haze-lines: Color restoration of underwater images," in Proc. British Mach. Vis. Conf. (BMVC) (London, UK, 2017).

[17] C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, and P. Bekaert, "Color balance and fusion for underwater image enhancement," IEEE Transactions on Image Processing, Vol. 27 (2018) No.1, p.379.

[18] J. Y. Chiang and Y.-C. Chen, "Underwater image enhancement by wavelength compensation and dehazing," IEEE Transactions on Image Processing, Vol. 21 (2012) No.4, p.1756.

[19] C. Li, J. Guo, and C. Guo, "Emerging from water: Underwater image color correction based on weakly supervised color transfer," IEEE Signal Processing Letters, Vol. 25 (2018) No.3, p.323.

[20] R. Schettini and S. Corchs, "Underwater image processing: state of the art of restoration and image enhancement methods," EURASIP Journal on Advances in Signal Processing, Vol. 2010 (2010) No.1, p.1.

[21] Y. T. Peng and P. C. Cosman, "Underwater image restoration based on image blurriness and light absorption," IEEE Transactions on Image Processing, Vol. 26 (2017) No.4, p.1579.

[22] H. Lu, Y. Li, S. Nakashima, and H. Kim, "Underwater image super-resolution by descattering and fusion," IEEE Access, Vol. 5 (2017), p.670.

[23] C. Li et al., "An underwater image enhancement benchmark dataset and beyond," IEEE Transactions on Image Processing, Vol. 29 (2020), p.4376.

[24] C. Li, S. Anwar, J. Hou, R. Cong, C. Guo, and W. Ren, "Underwater image enhancement via medium transmission-guided multi-color space embedding," IEEE Transactions on Image Processing, Vol. 30 (2021), p.4985.

[25] J. Li, K. A. Skinner, R. M. Eustice, and M. Johnson-Roberson, "WaterGAN: Unsupervised generative network to enable real-time color correction of monocular underwater images," IEEE Robotics and Automation Letters, Vol. 3 (2018) No.1, p.387.

[26] A. G. Howard et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," arXiv preprint arXiv:1704.04861 (2017).

[27] J. Kim, S. Lee, and S. S. Lee, "Real-time underwater image enhancement using embedded GPU," in Proc. MTS/IEEE OCEANS Conf. (Seattle, WA, USA, 2019), p.1.

[28] D. K. F. S. Rodrigues and A. C. P. L. F. de Carvalho, "Power-aware computing systems for autonomous underwater vehicles: A survey," ACM Computing Surveys, Vol. 54 (2021) No.1, p.1.

[29] P. Drews, E. Nascimento, F. Moraes, S. Botelho, and M. Campos, "Transmission estimation in underwater single images," in Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW) (Sydney, Australia, 2013), p.825.

[30] M. Stojanovic and J. Preisig, "Underwater acoustic communication channels: Propagation models and statistical characterization," IEEE Communications Magazine, Vol. 47 (2009) No.1, p.84.

[31] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. Int. Conf. Mach. Learn. (ICML) (Long Beach, CA, USA, 2019), p.6105.

[32] Y. Wang, J. Zhang, Y. Cao, and G. Wang, "A deep CNN method for underwater image enhancement," in Proc. IEEE Int. Conf. Image Process. (ICIP) (Abu Dhabi, United Arab Emirates, 2020), p.1.

[33] H. Lu, Y. Li, L. Zhang, and S. Serikawa, "Underwater image enhancement using guided trigonometric bilateral filter and fast automatic color correction," in Proc. IEEE Int. Conf. Image Process. (ICIP) (Phoenix, AZ, USA, 2016), p.1.

[34] N. Ma, X. Zhang, H. T. Zheng, and J. Sun, "ShuffleNet V2: Practical guidelines for efficient CNN architecture design," in Proc. Eur. Conf. Comput. Vis. (ECCV) (Munich, Germany, 2018), p.116.

[35] A. G. Howard et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW) (Honolulu, HI, USA, 2017), p.1.

[36] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) (Honolulu, HI, USA, 2017), p.1251.

[37] R. Liu, X. Fan, M. Zhu, M. Hou, and Z. Luo, "Real-time underwater image enhancement with Shallow-UWNet," IEEE Journal of Oceanic Engineering, Vol. 46 (2021) No.3, p.823.

[38] S. Zhao, B. Li, and Y. Xu, "EfficientUnderwater: A lightweight enhancement network based on EfficientNet," in Proc. Int. Conf. Artif. Intell. Appl. (ICAIA) (Shanghai, China, 2022), p.245.

[39] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An extremely efficient convolutional neural network for mobile devices," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) (Salt Lake City, UT, USA, 2018), p.6848.

[40] N. Ma, X. Zhang, H. T. Zheng, and J. Sun, "ShuffleNet V2: Practical guidelines for efficient CNN architecture design," arXiv preprint arXiv:1807.11164 (2018).

[41] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) (Las Vegas, NV, USA, 2016), p.770.

[42] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) (Salt Lake City, UT, USA, 2018), p.4510.

[43] Y. Wang, H. Liu, and L. P. Chau, "Underwater image enhancement using multi-scale dense generative adversarial network," IEEE Journal of Oceanic Engineering, Vol. 46 (2021) No.3, p.964.

[44] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) (Salt Lake City, UT, USA, 2018), p.7132.

[45] A. Howard et al., "Searching for MobileNetV3," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV) (Seoul, South Korea, 2019), p.1314.

[46] L. Xu, R. Wang, and J. Du, "Underwater image enhancement using SE-MobileNet," in Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW) (London, UK, 2020), p.1.

[47] Y. Chen, X. Dai, M. Liu, D. Chen, L. Yuan, and Z. Liu, "Dynamic convolution: Attention over convolution kernels," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) (Seattle, WA, USA, 2020), p.11030.

[48] Z. Guo, H. Li, and Z. Zhang, "Adaptive underwater image enhancement using dynamic convolution networks," IEEE Transactions on Circuits and Systems for Video Technology, Vol. 32 (2022) No.6, p.3875.

[49] P. Luo, J. Zhanglin, S. Zhang, M. Ren, and Z. Zhang, "Switchable normalization for learning-to-normalize deep representation," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 43 (2021) No.2, p.712.

[50] Z. Wang, J. Li, H. Wang, and H. Ling, "Dynamic routing for efficient network architecture," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS) (Vancouver, Canada, 2020), p.1.

[51] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) (Seattle, WA, USA, 2020), p.11531.

[52] H. Zhang, Y. Sun, and J. Liu, "ECA-UWNet: Efficient channel attention for underwater image enhancement," IEEE Access, Vol. 9 (2021), p.123456.

[53] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in Proc. Eur. Conf. Comput. Vis. (ECCV) (Munich, Germany, 2018), p.3.

[54] R. Li, J. Pan, and Z. He, "Lightweight spatial attention for real-time image enhancement," IEEE Transactions on Image Processing, Vol. 30 (2021), p.7898.

[55] Y. Wang, J. Guo, H. Li, and Q. Wu, "PA-URLNet: Pixel-aware lightweight underwater image restoration network," IEEE Transactions on Circuits and Systems for Video Technology, Vol. 33 (2023) No.4, p.1728.

[56] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) (Salt Lake City, UT, USA, 2018), p.7794.

[57] Z. Huang, X. Wang, L. Huang, C. Huang, Y. Wei, and W. Liu, "CCNet: Criss-cross attention for semantic segmentation," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV) (Seoul, South Korea, 2019), p.603.

[58] Y. Cao, J. Xu, S. Lin, F. Wei, and H. Hu, "GCNet: Non-local networks meet squeeze-excitation networks and beyond," in Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW) (Seoul, South Korea, 2019), p.1.

[59] Y. Chen, Y. Kalantidis, J. Li, S. Yan, and J. Feng, "A^2-Nets: Double attention networks," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS) (Montreal, Canada, 2018), p.350.

[60] H. Liu, K. Simonyan, and Y. Yang, "DARTS: Differentiable architecture search," in Proc. Int. Conf. Learn. Representations (ICLR) (New Orleans, LA, USA, 2019).

[61] Z. Zhang, H. Zheng, and J. Hong, "Neural architecture search for underwater image enhancement," IEEE Transactions on Geoscience and Remote Sensing, Vol. 60 (2022), p.1.

[62] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. Int. Conf. Mach. Learn. (ICML) (Long Beach, CA, USA, 2019), p.6105.

[63] L. Wang, X. Wang, and Y. Wang, "EfficientNet-based underwater image enhancement with attention mechanism," IEEE Journal of Oceanic Engineering, Vol. 47 (2022) No.2, p.512.

[64] H. Cai, L. Zhu, and S. Han, "ProxylessNAS: Direct neural architecture search on target task and hardware," in Proc. Int. Conf. Learn. Representations (ICLR) (New Orleans, LA, USA, 2019).

[65] B. Wu et al., "FBNet: Hardware-aware efficient convnet design via differentiable neural architecture search," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) (Long Beach, CA, USA, 2019), p.10726.

[66] J. Choi, Z. Wang, S. Venkataramani, P. I. J. Chuang, V. Srinivasan, and K. Gopalakrishnan, "PACT: Parameterized clipping activation for quantized neural networks," arXiv preprint arXiv:1805.06085 (2018).

[67] H. Pham, M. Y. Guan, B. Zoph, Q. V. Le, and J. Dean, "Efficient neural architecture search via parameter sharing," in Proc. Int. Conf. Mach. Learn. (ICML) (Stockholm, Sweden, 2018), p.4095.

[68] Z. Guo, X. Zhang, H. Mu, W. Heng, Z. Liu, and J. Sun, "Single path one-shot neural architecture search with uniform sampling," in Proc. Eur. Conf. Comput. Vis. (ECCV) (Glasgow, UK, 2020), p.544.

[69] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," arXiv preprint arXiv:1503.02531 (2015).

[70] J. Ba and R. Caruana, "Do deep nets really need to be deep?," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS) (Montreal, Canada, 2014), p.2654.

[71] T. Wang, R. Li, Z. Sun, and H. Wang, "Lightweight GAN for underwater image enhancement via knowledge distillation," in Proc. MTS/IEEE OCEANS Conf. (San Diego, CA, USA, 2021), p.1.

[72] S. Zagoruyko and N. Komodakis, "Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer," in Proc. Int. Conf. Learn. Representations (ICLR) (Toulon, France, 2017).

[73] A. Romero, N. Ballas, S. E. Kahou, A. Chassang, C. Gatta, and Y. Bengio, "FitNets: Hints for thin deep nets," in Proc. Int. Conf. Learn. Representations (ICLR) (San Diego, CA, USA, 2015).

[74] Y. Chen, N. Wang, and Z. Zhang, "DarkRank: Accelerating deep metric learning via cross sample similarities transfer," in Proc. AAAI Conf. Artif. Intell. (New Orleans, LA, USA, 2018), p.1.

[75] B. Peng, X. Jin, J. Liu, D. Li, Y. Wu, and J. Liu, "Correlation congruence for knowledge distillation," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV) (Seoul, South Korea, 2019), p.5006.

[76] H. Li, A. Kadav, I. Durdanovic, H. Samet, and H. P. Graf, "Pruning filters for efficient convnets," in Proc. Int. Conf. Learn. Representations (ICLR) (Toulon, France, 2017).

[77] Z. Liu, J. Li, Z. Shen, G. Huang, S. Yan, and C. Zhang, "Learning efficient convolutional networks through network slimming," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV) (Venice, Italy, 2017), p.2736.

[78] J. He, Y. Zhang, and S. Li, "Pruning underwater image enhancement networks for embedded deployment," IEEE Transactions on Circuits and Systems for Video Technology, Vol. 32 (2022) No.7, p.4321.

[79] S. Han, H. Mao, and W. J. Dally, "Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding," in Proc. Int. Conf. Learn. Representations (ICLR) (San Juan, Puerto Rico, 2016).

[80] J. Frankle and M. Carbin, "The lottery ticket hypothesis: Finding sparse, trainable neural networks," in Proc. Int. Conf. Learn. Representations (ICLR) (New Orleans, LA, USA, 2019).

[81] M. Lin, R. Ji, Y. Wang, Y. Zhang, B. Zhang, and Y. Tian, "HRank: Filter pruning using high-rank feature map," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) (Seattle, WA, USA, 2020), p.1529.

[82] R. Yu, A. Li, C. F. Chen, J. H. Lai, V. I. Morariu, and L. Davis, "NISP: Pruning networks using neuron importance score propagation," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) (Salt Lake City, UT, USA, 2018), p.9194.

[83] B. Jacob et al., "Quantization and training of neural networks for efficient integer-arithmetic-only inference," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) (Salt Lake City, UT, USA, 2018), p.2704.

[84] NVIDIA Corporation, "NVIDIA TensorRT: Programmable inference accelerator," [Online]. Available: https://developer.nvidia.com/tensorrt.

[85] K. Wang, Z. Liu, Y. Lin, J. Lin, and S. Han, "HAQ: Hardware-aware automated quantization with mixed precision," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) (Long Beach, CA, USA, 2019), p.8612.

[86] Y. Bhalgat, J. Lee, M. Nagel, T. Blankevoort, and N. Kwak, "LSQ+: Improving low-bit quantization through learnable offsets and better initialization," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW) (Seattle, WA, USA, 2020), p.1.

[87] Z. Dong, Z. Yao, A. Gholami, M. W. Mahoney, and K. Keutzer, "HAWQ: Hessian aware quantization of neural networks with mixed-precision," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV) (Seoul, South Korea, 2019), p.293.

[88] H. Wu, P. Judd, X. Zhang, M. Isaev, and P. Micikevicius, "Integer quantization for deep learning inference: Principles and empirical evaluation," arXiv preprint arXiv:2004.09602 (2020).

[89] M. Rastegari, V. Ordonez, J. Redmon, and A. Farhadi, "XNOR-Net: ImageNet classification using binary convolutional neural networks," in Proc. Eur. Conf. Comput. Vis. (ECCV) (Amsterdam, The Netherlands, 2016), p.525.

[90] G. Huang, S. Liu, L. Van der Maaten, and K. Q. Weinberger, "CondenseNet: An efficient densenet using learned group convolutions," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) (Salt Lake City, UT, USA, 2018), p.2752.

[91] L. Chen, H. Wang, and Y. Zhang, "Condense-UWNet: Efficient underwater image enhancement using learned group convolutions," IEEE Transactions on Geoscience and Remote Sensing, Vol. 60 (2022), p.1.

[92] J. Ballas, L. Yao, C. Pal, and A. Courville, "Delving deeper into convolutional networks for learning video representations," in Proc. Int. Conf. Learn. Representations (ICLR) (San Juan, Puerto Rico, 2016).

[93] S. Li, W. Li, C. Cook, C. Zhu, and Y. Gao, "Independently recurrent neural network (IndRNN): Building a longer and deeper RNN," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) (Salt Lake City, UT, USA, 2018), p.5457.

[94] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," in Proc. Int. Conf. Learn. Representations (ICLR) (Virtual Event, 2021).

[95] S. Mehta and M. Rastegari, "MobileViT: Light-weight, general-purpose, and mobile-friendly vision transformer," in Proc. Int. Conf. Learn. Representations (ICLR) (Virtual Event, 2022).

[96] B. Graham, A. El-Nouby, H. Touvron, P. Stock, A. Joulin, and H. Jégou, "LeViT: A vision transformer in ConvNet's clothing for faster inference," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV) (Montreal, Canada, 2021), p.12259.

[97] S. Yu, T. Chen, J. Shen, H. Yuan, J. Tan, and S. Liu, "PoolFormer: Exploring pooling for vision transformer," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) (New Orleans, LA, USA, 2022), p.8877.

[98] X. Li, H. Wang, and L. Zhang, "Transformer-based lightweight underwater image enhancement," IEEE Transactions on Image Processing, Vol. 32 (2023), p.1234.

[99] M. J. Islam, Y. Xia, and J. Sattar, "Fast underwater image enhancement for improved visual perception," IEEE Robotics and Automation Letters, Vol. 5 (2020) No.2, p.3227.

[100] M. J. Islam, C. Edge, Y. Xiao, P. Luo, M. Mehtaz, and C. Morse, "Semantic segmentation of underwater imagery: Dataset and benchmark," in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS) (Las Vegas, NV, USA, 2020), p.1769.

[101] X. Fu, P. Zhuang, Y. Huang, Y. Liao, X. P. Zhang, and X. Ding, "A retinex-based enhancing approach for single underwater image," in Proc. IEEE Int. Conf. Image Process. (ICIP) (Paris, France, 2014), p.4572.

[102] B. L. McGlamery, "A computer model for underwater camera systems," in Proc. SPIE Ocean Optics VI (Monterey, CA, USA, 1980), p.221.

[103] D. Akkaynak and T. Treibitz, "Sea-thru: A method for removing water from underwater images," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) (Long Beach, CA, USA, 2019), p.1682.

[104] H. Blasinski, J. Farrell, and T. Lian, "Underwater image formation model and its application to restoration and enhancement," in Proc. IEEE Int. Conf. Comput. Photogr. (ICCP) (Evanston, IL, USA, 2018), p.1.

[105] C. Fabbri, M. J. Islam, and J. Sattar, "Enhancing underwater imagery using generative adversarial networks," in Proc. IEEE Int. Conf. Robot. Autom. (ICRA) (Brisbane, Australia, 2018), p.7159.

[106] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) (Las Vegas, NV, USA, 2016), p.2414.

[107] X. Huang and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV) (Venice, Italy, 2017), p.1501.

[108] Y. Wang, J. Zhang, Y. Cao, and Z. Wang, "A deep CNN method for underwater image enhancement based on physical model and deep learning," IEEE Transactions on Circuits and Systems for Video Technology, Vol. 31 (2021) No.8, p.3078.

[109] D. Berman, T. Treibitz, and S. Avidan, "Non-local image dehazing," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) (Las Vegas, NV, USA, 2016), p.1674.

[110] X. Liu, Z. Gao, and R. Wu, "Underwater image enhancement with scene depth guidance," IEEE Transactions on Image Processing, Vol. 30 (2021), p.6545.

[111] C. Ledig et al., "Photo-realistic single image super-resolution using a generative adversarial network," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) (Honolulu, HI, USA, 2017), p.4681.

[112] P. Charbonnier, L. Blanc-Feraud, G. Aubert, and M. Barlaud, "Two deterministic half-quadratic regularization algorithms for computed imaging," in Proc. IEEE Int. Conf. Image Process. (ICIP) (Austin, TX, USA, 1994), p.168.

[113] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Transactions on Image Processing, Vol. 13 (2004) No.4, p.600.

[114] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in Proc. Asilomar Conf. Signals, Syst. Comput. (Pacific Grove, CA, USA, 2003), p.1398.

[115] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in Proc. Eur. Conf. Comput. Vis. (ECCV) (Amsterdam, The Netherlands, 2016), p.694.

[116] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) (Las Vegas, NV, USA, 2016), p.2414.

[117] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) (Salt Lake City, UT, USA, 2018), p.586.

[118] G. D. Finlayson, S. D. Hordley, and P. M. Hubel, "Color by correlation: A simple, unifying framework for color constancy," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 23 (2001) No.11, p.1209.

[119] R. C. Gonzalez and R. E. Woods, Digital Image Processing, 4th ed. (Pearson, New York 2018).

[120] I. Goodfellow et al., "Generative adversarial nets," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS) (Montreal, Canada, 2014), p.2672.

[121] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV) (Venice, Italy, 2017), p.2794.

[122] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of Wasserstein GANs," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS) (Long Beach, CA, USA, 2017), p.5767.

[123] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation-based noise removal algorithms," Physica D: Nonlinear Phenomena, Vol. 60 (1992) No.1-4, p.259.

[124] C. Li, S. Anwar, and F. Porikli, "Underwater scene prior inspired deep underwater image and video enhancement," Pattern Recognition, Vol. 98 (2020), p.107038.

[125] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," in Proc. Int. Conf. Learn. Representations (ICLR) (Vancouver, Canada, 2018).

[126] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," in Proc. Int. Conf. Mach. Learn. (ICML) (Montreal, Canada, 2009), p.41.

[127] A. Saxena, S. H. Chung, and A. Y. Ng, "Learning depth from single monocular images," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS) (Vancouver, Canada, 2006), p.1.

[128] T. Chen, S. Kornblith, K. Swersky, M. Norouzi, and G. Hinton, "Big self-supervised models are strong semi-supervised learners," arXiv preprint arXiv:2006.10029 (2020).

[129] R. Cong, H. Zhao, C. Li, J. Zhang, F. Zheng, and Y. Zhao, "A Comprehensive Survey on Underwater Image Enhancement Based on Deep Learning," arXiv preprint arXiv:2405.19684 (2024).

[130] T. Cao, J. Yu, and Y. Liu, "DNnet: A lightweight network for real-time 4K underwater image enhancement using dynamic range and average normalization," Expert Systems with Applications, Vol. 251 (2025), p.126561.

[131] A. Naik, A. Swarnakar, and K. Mittal, "Shallow-UWNet: Compressed model for underwater image enhancement," arXiv preprint arXiv:2101.02073 (2021).

[132] Z. Yang, H. Xu, J. Wang, and C. Wang, "LU2Net: A Lightweight Network for Real-time Underwater Image Enhancement," arXiv preprint arXiv:2406.14973 (2024).

[133] Y. Liu, X. Yang, and L. Wang, "A Novel Lightweight Model for Underwater Image Enhancement," Sensors, Vol. 24 (2024) No.10, p.3070.

[134] W. Li, H. Li, and Z. Zhang, "Lightweight underwater image adaptive enhancement based on zero-reference parameter estimation network," Frontiers in Marine Science, Vol. 11 (2024), p.1378817.

[135] J. Xie, C. Gao, and Y. Wang, "Underwater image enhancement based on zero-shot learning and underwater scene prior," Applied Intelligence, Vol. 53 (2023) No.12, p.15045.