Xu, J. et al. (Eds.) (2026). Proceedings of the 34th International Conference on Computers in Education.
Asia-Pacific Society for Computers in Education

Designing a Multi-Agent System for AIED Ethics Learning

Jie XU[a,b], Wenli CHEN[b*], Yuan-Hao JIANG[b,c], Xuanyu CHEN[b], Hong Poh LIANG[b] & Yu LU[d]
[a]*College of Education, Zhejiang University, Zhejiang, China*
[b]*National Institute of Education, Nanyang Technological University, Singapore*
[c]*Shanghai Institute of Artificial Intelligence for Education, East China Normal University, Shanghai, China*
[d]*School of Educational Technology, Beijing Normal University, Beijing, China*
*Corresponing author: wenli.chen@nie.edu.sg

Abstract: This study designed and developed a multi-agent system for teacher learning in AIED ethics, grounded in experiential learning theory. The system structures teachers' learning as a cyclical process of concrete experience, reflective observation, abstract conceptualization, and active experimentation through the collaboration of five agents. This study proposed a contextualized, interactive, and reflective approach to AIED ethics learning and offers practical implications for the design and application of multi-agent systems in teacher professional learning.

Keywords: AIED ethics, teachers, multi-agent system, experiential learning theory

1. Introduction

The rapid development of artificial intelligence in education (AIED) is reshaping teaching, learning, and educational management. On one hand, AI can support personalized learning, enhance teachers' efficiency, and automate certain administrative and procedural tasks (Ocak & Caskurlu, 2026). On the other hand, the growing integration of emerging AI technologies into educational practice has raised concerns, regarding data privacy, algorithmic bias, academic integrity, and learner agency (Abuadas & Albikawi, 2026; Dharmapatni et al., 2026). Given their roles as key decision-makers in AIED practice, teachers' awareness and knowledge of AIED ethics have attracted increasing attention. UNESCO's (2024) *AI Competency Framework for Teachers* identified AI ethics as an important component of teachers' AI competencies, while Ministry of Education of Singapore (2025) highlighted Agency, Inclusivity, Fairness, and Safety (AIFS) as four key dimensions of ethical AI use in educational practice. Teachers' professional preparedness extends beyond technical competence to encompass the ability to identify, analyze, and address ethical issues arising from AI in specific educational contexts (Firdian, 2026).

Recent research has focused on the ethical issues associated with teachers' use of AI in educational practice. Ho & Lee (2025) found that teachers' understanding of AI ethics evolved from an initial focus on functionality and efficiency to recognition of ethical concerns, followed by reflection, reassessment, and adjustments in practice. Teachers gradually became aware of the ethical implications of AI use as issues such as student over-reliance, content bias, and privacy breaches emerged in practice. Similarly, Dharmapatni et al. (2026) found that although secondary school English teachers reported relatively high self-efficacy in using AI, they still struggled to identify AI misuse and assess the authenticity of students' writing. Ocak & Caskurlu (2026) likewise found that teachers with limited foundational technological knowledge often struggled to understand how human decisions influence AI systems and their outcomes. These studies suggest that the ethical issues teachers encounter in AI use are complex and context-dependent, and that their ethical judgments develop through engagement with specific issues and subsequent reflection.

Existing research on AIED and teacher professional development has provided limited support for teachers in addressing ethical issues in practice. A systematic review by Mouta et

al. (2024) found that, although some schools had introduced relevant training, ethical consideration remained insufficiently addressed. Ma et al. (2025) reported that most secondary school teachers lacked confidence in teaching AI ethics due to insufficient systematic ethics education and professional training. They also found that one-off, technology-oriented training was inadequate to address the complex demands of AIED ethics. Related research (e.g., Mouta et al., 2025) suggests that teachers need more opportunities to develop their AIED ethical awareness through practical activities, hands-on experience, and professional dialogue. Therefore, teachers need contextualized, interactive, and reflective learning opportunities to identify ethical issues, consider multiple perspectives in authentic educational contexts, and develop ethical judgments that can inform future practice.

AI-based learning environments offer a promising way to provide such support through scenario generation, and adaptive feedback. A multi-agent system (MAS) consists of multiple autonomous agents that can collaborate or compete with one another to jointly accomplish complex tasks (Janbi et al., 2023). Compared with approaches in which a single agent primarily delivers knowledge or feedback to learners, multi-agent systems can assign different agents the roles of teachers, peers, mentors, or other stakeholders, thereby simulating interaction processes that more closely resemble authentic educational environments. In recent years, advances in large language models have further strengthened the interaction and reasoning capabilities of MAS, enabling agents to communicate dynamically through natural language and to perform functions such as contextual understanding, role-playing, and feedback generation (Jiang et al., 2025). Ethical issues in AIED often involve multiple values, while also engaging diverse stakeholders such as teachers, students, school leaders, and developers. As a result, such issues rarely have a single standard answer. By allowing different agents to represent different stakeholders, MAS can transform abstract ethical principles into concrete ethical scenarios, enabling teachers to experience conflicting perspectives, make ethical judgments, and reflect on their decisions.

However, despite growing attention to AIED ethics, how MAS can be designed to support teachers in understanding and addressing complex ethical situations, as well as their actual effects in such learning contexts, remains insufficiently examined (Guo et al., 2024). Therefore, this study further investigates how experiential learning theory can be integrated with multi-agent role interaction, contextualized decision-making, feedback, and reflection to support teachers' learning of AIED ethics.

2. Design and Implementation of the Multi-Agent System

2.1 design framework

The multi-agent learning design is informed by Kolb's (1984) experiential learning theory (ELT), which has been widely applied in education. ELT conceptualizes learning as a continuous four-stage cycle comprising Concrete Experience (CE), Reflective Observation (RO), Abstract Conceptualization (AC), and Active Experimentation (AE) (Kolb, 1984). Learners move from concrete experience to reflection, abstract conceptualization, and the application of newly developed understandings in new situations, forming a continuous learning cycle. A key emphasis of ELT is the transformation of experience and reflection into understanding and practice, which aligns well with teachers' learning about AIED ethics. Teachers need not only to understand ethical principles but also to identify and reflect on ethical issues in specific contexts, develop broader ethical understandings, and translate them into judgments and actions for future educational practice.

Grounded in experiential learning theory, this study developed a multi-agent-supported learning cycle for teachers' learning of AIED ethics. The CE stage emphasizes learning through authentic or realistic situations, enabling teachers to perceive and form initial judgments about AIED-related ethical issues. The system includes a Scenario Generation Agent that provides teachers with practice-oriented ethical scenarios to elicit their initial responses. The RO stage requires learners to revisit prior experiences and examine relevant issues and implications from multiple perspectives. Accordingly, the study designed

an Ethical Analysis Agent and a Socratic Questioning Agent (Etori & Gini, 2024). The Ethical Analysis Agent diagnoses teachers' responses, while the Socratic Questioning Agent uses reflective questions to deepen their examination of their own judgments. As reflection deepens, the AC stage requires teachers to move beyond specific cases and develop broader ethical understandings, criteria for judgment, and principles for action. Accordingly, there is an Ethical Conceptualization Agent to help teachers synthesize and generalize their context-specific reflections. Finally, the AE stage emphasizes applying newly developed understandings to future practice. Accordingly, the system guides teachers to translate their ethical understanding into transferable rules or action plans. To ensure alignment between the learning process and its ethical goals, this study designed an Evidence and Consistency Monitoring Agent to monitor the quality and consistency of outputs throughout the interaction.

2.2 Ethical Scenario Knowledge Base and Design Goals

Guidance on AI ethics in education is spread across many documents issued by international organizations and national education authorities, which differ in scope, in level of detail, and at times in the positions they take. When a large language model answers ethical questions without reference to such documents, it can produce advice that sounds reasonable but cannot be traced to a source. A knowledge base of ethical scenarios was therefore built for the agents to draw on. Taking the Singapore Ministry of Education's AIED Ethics Framework and its four principles as a starting point, official documents were then collected from UNESCO, the OECD, the European Commission, and the U.S. Department of Education.

Five design goals were drawn from the knowledge base, each linked to a stage of Kolb's (1984) experiential learning cycle (see Table 1). DG1 to DG3 and DG5 correspond to successive stages of the cycle. DG4 applies throughout because every entry carries a chapter and page reference; the agents can be required to name the passage on which a claim rests. This reduces the risk of fabricated content and leaves authority for ethical claims with the source document rather than the model, in line with the human-in-the-loop principle that the documents themselves set out.

Table 1. *Design Goals, Their Grounding in the Knowledge Base, and Learning Stages*

Design goal	**Grounding in the knowledge base**	**Stage supported**
DG1. Authenticity of experience	16 workflow-indexed cases documenting real practice and its attendant risks	Concrete experience
DG2. Structured interrogation	Guiding questions and the "possible risks" columns of the use-case tables	Reflective observation
DG3. Multi-perspectivity	11 principles from four issuing bodies, with inter-source tensions preserved	Abstract conceptualization
DG4. Traceability and oversight	Verbatim excerpts with chapter- and page-level provenance	Cross-cutting
DG5. Calibrated progression	Acquire-Deepen-Create levels; 22 AI literacy competencies	Active experimentation

2.3 Multi-Agent Architecture and Workflow

As shown in Fig. 1, the system workflow consists of two main phases. First, during the Scenario Setup phase, teachers select or enter an educational context and an ethical issue they wish to explore. Based on this input, the Scenario Generation Agent generates a specific AIED ethics scenario together with an open-ended question, after which the teacher provides an initial judgment on the case. Second, during the Interactive Rounds phase, the Ethical Analysis Agent examines the ethical dimensions already addressed in the teacher's response as well as those that may have been overlooked and identifies the focus for subsequent interaction. Based on this analysis, the Socratic Questioning Agent generates adaptive, reflective questions that prompt teachers to consider additional stakeholders, potential risks, value conflicts, and ethical issues that have not yet been sufficiently examined.

After the teacher responds, the Ethical Conceptualization Agent provides a stage-based summary of the teacher's reasoning and further supports the transformation of case-specific judgments into more general ethical principles or transferable guidelines for action. If important ethical dimensions remain insufficiently explored, the system proceeds to another round of interaction. Once the teacher has developed a relatively comprehensive ethical judgment, the system generates a summary and concludes the current learning process. Meanwhile, the Evidence and Consistency Monitoring Agent check key outputs for deviations from the scenario, logical inconsistencies, or premature disclosure of answers, but does not interact directly with the teacher. In this way, the multi-agent coordination mechanism illustrated in Fig. 1 supports teachers in progressing through a learning process of "concrete situation–reflection–conceptualization–transfer to action," rather than having AI directly provide predetermined ethical conclusions.

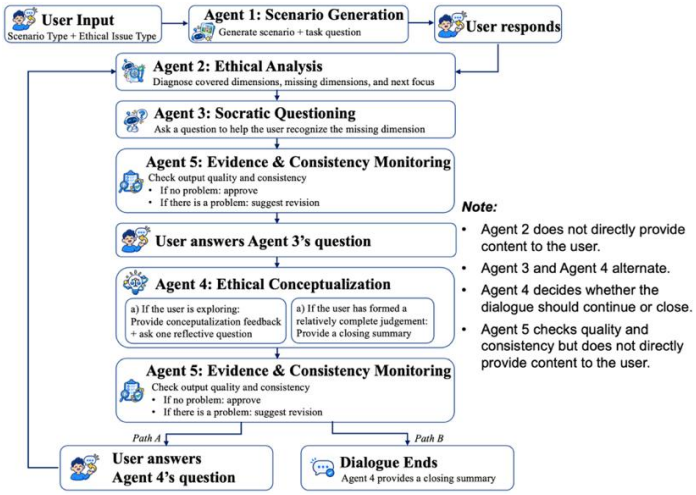

Figure 1. Multi-agent Architecture and Workflow for AIED Ethics Learning.

2.4 System Implementation

The multi-agent system was implemented in the Leadership in Artificial Intelligence for Education course at the National Institute of Education (NIE), Nanyang Technological University, and was integrated as part of the course activities on AIED ethics. The course is designed for educational leaders and aims to help learners understand both the potential and limitations of AI in education, while developing their capacity to select, implement, and lead AI applications in school contexts. Its content explicitly addresses ethical issues in AIED, including privacy, bias, transparency, and explainability, as well as trade-offs between system functionality and ethical use. This provided an instructional context that is closely aligned with the design goals of the proposed system.

During the course activity, teachers first accessed the system and selected or entered an educational context and an ethical issue they wished to explore. The system then generated a corresponding AIED ethics scenario, and teachers were asked to make an initial judgment about the issue presented in the case and respond to reflective questions over multiple rounds of dialogue. As the interaction progressed, the system dynamically adjusted subsequent questions based on teachers' responses and, at the end of the learning process, summarized the major ethical concepts and key insights developed during the discussion. The course implementation was approved by the relevant institutional ethics review process. The data analysis of the system evaluation is still ongoing.

3 Conclusion

This study designed and implemented a multi-agent system for AIED ethics learning based on experiential learning theory. Through scenario generation, Socratic questioning, and ethical

conceptualization, the system guides teachers to reflect on concrete cases and gradually develop transferable ethical understanding. Unlike conversational systems that directly provide ethical knowledge or predetermined answers, the proposed system positions AI as a cognitive scaffold for teachers' ethical learning while retaining active experimentation as a teacher-driven process, thereby emphasizing teachers' agency in ethical judgment and action-oriented decision-making. Future research will combine pre- and post-tests, ethical scenario tasks, system interaction logs, and interviews to further examine the system's effects on teachers' identification of ethical issues, ethical reasoning, and action-oriented decision-making, as well as the roles of different agent-based scaffolds in the learning process.

Acknowledgements

This study was funded by INEI 3/25 CW supported by International Network of Educational Institutes. This study was approved by the NTU Institutional Review Board (IRB-2026-486).

References

Abuadas, M., & Albikawi, Z. (2026). AI ethical awareness and academic integrity in higher education: Development and validation of a new scale. *Ethics & Behavior*, *36*(2), 125–142. https://doi.org/10.1080/10508422.2025.2511336

Dharmapatni, P. N. P. V., Santosa, M. H., & Paramartha, A. A. G. Y. (2026). Imbalance in teacher readiness between technological self-efficacy and ethical awareness in the era of AI humanification. *Jurnal Studi Guru Dan Pembelajaran*, *9*(1), 563–757.

Etori, N. A., & Gini, M. (2024). WisCompanion: Integrating the Socratic Method with ChatGPT-Based AI for Enhanced Explainability in Emotional Support for Older Adults. In H. Degen & S. Ntoa (Eds.), *Artificial Intelligence in HCI* (Vol. 14734, pp. 179–198). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-60606-9_11

Firdian, F. (2026). Preparing Educators for the Artificial Intelligence Era: Digital Competencies, Pedagogical Readiness, and Ethical Challenges in Education. *International Journal of Education and Vocational Studies*, *1*(1), 33–39. https://doi.org/10.62712/ijevs.v1i1.5

Guo, T., Chen, X., Wang, Y., Chang, R., Pei, S., Chawla, N. V., Wiest, O., & Zhang, X. (2024). *Large Language Model based Multi-Agents: A Survey of Progress and Challenges* (arXiv:2402.01680; Version 2). arXiv. http://arxiv.org/abs/2402.01680

Janbi, N., Katib, I., & Mehmood, R. (2023). Distributed artificial intelligence: Taxonomy, review, framework, and reference architecture. *Intelligent Systems with Applications*, *18*, 200231. https://doi.org/10.1016/j.iswa.2023.200231

Jiang, Y.-H., Lu, Y., Dai, L., Wang, J., Li, R., & Jiang, B. (2025, December). Agentic workflow for education: Concepts and applications. *Proceedings of the 33rd International Conference on Computers in Education*. https://library.apsce.net/index.php/ICCE/article/view/6061

Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development* (No. 1; Vol. 1). Prentice-hall Englewood Cliffs, NJ.

Ma, M., Ng, D. T. K., Liu, Z., Lin, J., & Wong, G. K. W. (2025). Why don't teachers teach AI ethics? Understanding teachers' beliefs and intentions in Chinese AI curriculum implementation through the theory of planned behaviour. *Computers and Education: Artificial Intelligence*, *9*, 100518. https://doi.org/10.1016/j.caeai.2025.100518

Ministry of Education of Singapore. (2025). *AI educational ethics framework*. https://www.learning.moe.edu.sg/ai-in-sls/responsible-ai/ai-in-education-framework/

Mouta, A., Pinto-Llorente, A. M., & Torrecilla-Sánchez, E. M. (2024). Uncovering Blind Spots in Education Ethics: Insights from a Systematic Literature Review on Artificial Intelligence in Education. *International Journal of Artificial Intelligence in Education*, *34*(3), 1166–1205. https://doi.org/10.1007/s40593-023-00384-9

Mouta, A., Torrecilla-Sánchez, E. M., & Pinto-Llorente, A. M. (2025). Comprehensive professional learning for teacher agency in addressing ethical challenges of AIED: Insights from educational design research. *Education and Information Technologies*, *30*(3), 3343–3387.

Ocak, C., & Caskurlu, S. (2026). Unpacking ethics-domain of intelligent-TPACK scale in relation to in-service teachers' trust and distrust. *Computers and Education Open*, *10*, 100321.

UNESCO. (2024). *AI competency framework for teachers*. https://doi.org/10.54675/ZJTE2084