MIPTS: A Multimodal Physics Tutoring System Synergizing Hybrid RAG and Autonomous Agents

Xiangyu Yang[1a*], Can He[2b], Meng Zhang[1c]

[1]Institute of Information and Computer, Wuhan College of Arts and Science, Wuhan, China
[2]Wuhan College of Arts and Science Medical College, Wuhan, China

[a] yangxy961103@gmail.com, [b] Hecan050724@outlook.com,
[c] zhangmeng1995edu@163.com

Abstract. Large Language Models (LLMs) have shown promise in educational applications, yet their direct use in physics tutoring is hindered by hallucinations, unstable symbolic computation, and insufficient modeling of physics knowledge structures. To address these limitations, this paper proposes a Multimodal Intelligent Physics Tutoring System (MIPTS) based on a Hybrid Retrieval-Augmented Generation (RAG) framework. The system integrates knowledge graphs, deep document understanding, and autonomous agents to support structured reasoning and teaching-oriented feedback in physics problem solving. An intention-driven dual-channel architecture separates latent tool-augmented reasoning from low-latency Socratic guidance, improving both reliability and instructional interpretability. Case studies on six university-level physics problems demonstrate that MIPTS achieves better physical consistency, reasoning transparency, and pedagogical rigor than general-purpose LLM-based systems.

Keywords: Multimodal Physics Tutoring; Hybrid RAG; Knowledge Graphs; Autonomous Agents; Symbolic Reasoning; Physics Education.

1 Introduction

In recent years, the rapid development of generative artificial intelligence represented by Large Language Models (LLMs) has been profoundly transforming the landscape of the education sector [1, 2]. AI Agents integrate the robust language modeling, cross-domain generalization, and natural interaction capabilities of LLMs, and are regarded as the pivotal technology for achieving large-scale personalized education and addressing the "2 Sigma Problem" [3].

In STEM education, artificial intelligence has been applied to implement Socratic tutoring and real-time feedback, which has been proven to enhance student engagement and conceptual comprehension [4]. Nevertheless, the direct application of general-purpose LLMs in physics education still faces challenges. Physics teaching involves abstract content such as rigorous reasoning and symbolic computation, which general-purpose models often struggle to handle competently.

I. A. Khan et al. (eds.), *Proceedings of the 2026 5th International Conference on Educational Innovation and Multimedia Technology (EIMT 2026)*, Atlantis Highlights in Social Sciences, Education and Humanities 51,
https://doi.org/10.2991/978-94-6239-691-3_14

Studies have indicated that general-purpose LLMs exhibit three major flaws in solving physics problems: first, they still generate factual hallucinations during complex reasoning processes, leading to task failure [5]; second, relying on probabilistic prediction, they lack stable computing capabilities [6], resulting in significant performance variations across different models and a high propensity for errors in physical explanations or mathematical derivations; third, they cannot explicitly model the structure of physical knowledge, making it difficult to constrain conceptual hierarchies and causal logic, which limits their application in physics education. This single-paradigm approach, devoid of constraints from authentic physical knowledge, fails to support the deep integration of artificial intelligence and teaching, nor can it meet the collaborative educational demands of multi-role, multi-stage, and multi-strategy personalized teaching [7].

To alleviate these issues, the Retrieval-Augmented Generation (RAG) framework has been adopted in educational intelligent systems, but traditional methods primarily rely on text-only vector retrieval [8]. Physical knowledge is highly dependent on a multimodal structured system encompassing formulas, diagrams, and other elements, rendering traditional RAG ineffective in physics education scenarios [9]. Knowledge Graphs (KGs) possess the advantage of explicitly representing knowledge structures and reasoning constraints, and are being integrated into RAG to compensate for its shortcomings in processing structured knowledge. Frameworks such as LightRAG, RagAnything, and RAGFlow have incorporated multimodal knowledge and knowledge graph technologies [10, 11]. However, how to achieve synergistic enhancement between knowledge graph retrieval and the distributed representations of LLMs remains an unsolved cutting-edge challenge.

In another research direction, intelligent tutoring systems based on multi-agent architecture have emerged as a research hotspot in education [12]. By constructing AI Agents with distinct functions to collaborate, these systems can more realistically simulate multi-role teaching mechanisms, support the provision of personalized services for students, and promote the automation and adaptive regulation of teaching processes. Research has shown that integrating LLM-based agents with memory, tool-use, and planning capabilities can deliver higher-quality instructional guidance and personalized support to both teachers and students.

Therefore, this paper proposes a **Multimodal Intelligent Physics Tutoring System** (MIPTS) that integrates knowledge graphs, deep document understanding, and multi-agent mechanisms. Unlike existing methods, the core objective of this study is not merely to improve the accuracy of problem-solving, but to construct a systematic intelligent tutoring framework characterized by knowledge consistency, reasoning interpretability, and adaptive teaching strategies in physics teaching scenarios.

2 System Architecture

To tackle the core challenges in physics teaching, including obstacles in multimodal understanding, instability in symbolic computation, and the absence of systematic teaching strategies, the Multimodal Intelligent Physics Teaching System (MIPTS) is

built upon the Dify [13] orchestration framework, adopting a closed-loop architecture of "Perception-Decision-Execution-Evolution". This section elaborates on the hierarchical architecture of the system and the collaborative workflow among its various functional modules in detail.

2.1 Overall Architecture and Collaborative Paradigm

As shown in Figure 1, MIPTS constructs a "Student-AI-Teacher" triangular trust paradigm with the Adaptive Physics Reasoning Engine as the core and multi-agent collaboration. This architecture consists of four key functional layers, which realize a closed-loop teaching process through the interaction of data flow and control flow.

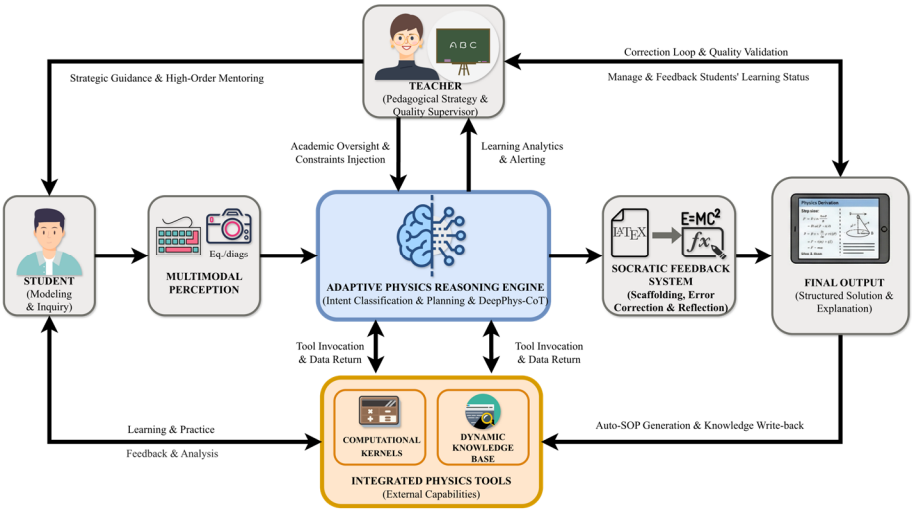

Fig. 1. The "Student-AI-Teacher" Collaborative Paradigm. The framework establishes a closed-loop synergy: (1) The Teacher acts as the pedagogical architect, injecting strategy and oversight. (2) The Adaptive Reasoning Engine (Center) performs intent classification and latent reasoning (DeepPhys-CoT), invoking external tools for verification. (3) The Socratic Feedback System delivers structured scaffolding to the Student. (4) Crucially, a Metacognitive Loop (Bottom-Right) abstracts successful problem-solving patterns into Standard Operating Procedures (SOPs), writing them back to the dynamic knowledge base for system evolution.

Serving as the system entry point, the multimodal perception module is responsible for processing the unstructured data input by students (including natural language texts, handwritten formula images, and physics diagrams), converting it into processable structured data, and conducting preliminary analysis. It leverages visual encoders and text parsers to transform multimodal inputs into unified structured semantic representations, which provide aligned contextual information for downstream reasoning. This effectively addresses the limitation that traditional text-based models are incapable of interpreting physics diagrams.

The Core Reasoning & Decision Layer serves as the "brain" of the system, which is composed of an Intent Classifier and a reasoning engine. The reasoning engine incorporates the DeepPhys-CoT (Deep Physics Chain-of-Thought) mechanism, which not only takes charge of planning problem-solving pathways but also acts as a central controller to dynamically schedule external tools (e.g., symbolic computation kernels) and knowledge base retrieval modules, ensuring that the reasoning process conforms to the causal logic of physics.

To reflect its pedagogical attributes, the system incorporates a Teacher Agent as a supervisor. Instead of directly engaging in problem-solving, it injects constraints into the reasoning engine based on the syllabus (a mechanism referred to as Constraints Injection), such as *"prohibiting the direct provision of answers"* or *"mandating dimensional analysis"*. Meanwhile, the Socratic Feedback System generates guiding follow-up questions based on reasoning outcomes rather than declarative answers, delivering interactive feedback to students and facilitating their self-adjustment throughout the learning process. Together, these components constitute the Pedagogical Supervision & Feedback Layer.

Positioned at the bottom layer of the architecture, the Procedural Memory module constitutes the system's long-term memory. It leverages an SOP Extractor to automatically abstract Standard Operating Procedures (SOPs) from successful problem-solving cases, and writes these "cognitive accumulations" back to the dynamic knowledge base, thereby enabling the system's self-iteration and evolution.

2.2 Intention-Driven Dual-Channel Workflow

The system operation process is illustrated in Figure 2, which adopts an intention-driven dual-track mechanism and dynamically switches between the concept learning pathway and the problem-solving pathway based on the nature of students' requests.

The system uses an intention classifier to determine whether a student's request is primarily for concept comprehension or targeted at specific problem-solving. Different judgment results will trigger distinct processing pathways and feedback strategies.

Track I (Implicit Reasoning Mode via Shadow Solver): When the Intent Classifier detects a new physics problem, the Shadow Solver (A customized scientific computing sandbox was developed based on the Dify platform, whose native sandbox only supports Python basic libraries.) is activated in Track B. This is a high-compute agent operating in the latent space, characterized by a "black-box reasoning, white-box verification" workflow:

1.Tool-Augmented Reasoning: The Shadow Solver invokes Python/SymPy for symbolic derivation or numerical tools for regression, ensuring the determinism of each mathematical transformation.

2. **Hybrid Retrieval Enhancement**: Before reasoning commences and when reasoning encounters obstacles, it triggers the **Hybrid Retrieval** module to simultaneously retrieve textbook texts from the vector database (we have built a repository of multiple physics textbooks including *University Physics*, *Analytical Mechanics*, *Quantum Mechanics*, and *Electrodynamics* via RAGFlow) and entity relationships from the knowledge graph (an intelligent knowledge graph has been constructed by combining

LightRAG, RagAnything, and large language models). This enables the acquisition of precise definitions of physical laws.

3. Full-Solution Caching: The generated ground truth is not output directly to the student but is structured and written into the global Solution Cache. This design provides a "standard reference" for subsequent pedagogical guidance while effectively preventing direct answer leakage.

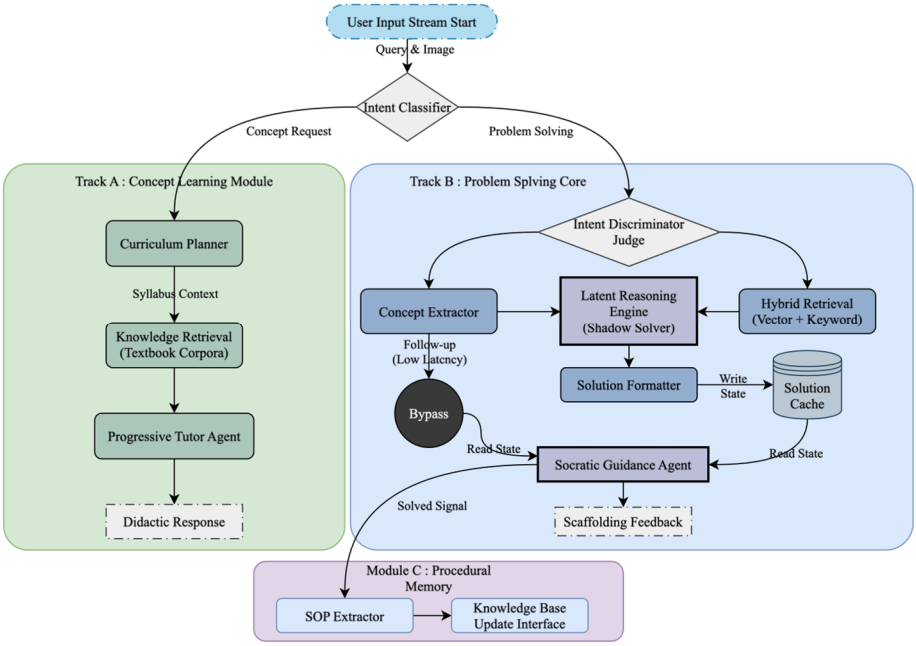

Fig. 2. The technical architecture of MIPTS. The system is structured into three coordinated modules: (1) Track A: Concept Learning Module (Left, green flow), which handles curriculum planning and textbook knowledge retrieval for conceptual inquiries. (2) Track B: Problem Solving Core (Right, blue flow), which executes the "High-Compute" latent reasoning via the Shadow Solver and utilizes a "Low-Latency" bypass mechanism for follow-up Socratic scaffolding. (3) Module C: Procedural Memory (Bottom, purple flow), which acts as a metacognitive loop to abstract successful problem-solving patterns into SOPs for system evolution.

Track II (Heuristic Bypass for Socratic Guidance): When students raise follow-up questions or request hints regarding existing problems, the system skips the shadow solver and directly switches to the low-latency Socratic Guidance Agent. This agent retrieves pre-generated problem-solving states from the Solution Cache with a time complexity of $O(1)$. Based on the retrieved "standard pathway" and current conversational context, it adopts a Scaffolding Strategy to generate rhetorical questions or hints (e.g., "Have you considered conducting a force analysis on this object?"), thereby guiding students to construct knowledge independently.

Both Track I and Track II ultimately lead the system to the Socratic Guidance Agent, which then interacts with students using the scaffolding strategy. When the system detects that a problem has been successfully solved, the problem-solving process is sent to the procedural memory module, where the problem-solving process extractor converts it into structured knowledge and writes it back to the knowledge base, thereby supporting the continuous optimization of the system in subsequent teaching activities. Meanwhile, the system dynamically records students' learning states as knowledge flows in the system knowledge base, achieving personalized learner profile descriptions for different students. As students increase their interactions with MIPTS, the system gains a deeper understanding of their learning progress.

3 Case Studies and System Comparison

To verify the effectiveness of MIPTS in physics teaching scenarios, particularly its advantages in multimodal understanding, mathematical symbolic reasoning, and teaching process transparency, we designed a series of comparative experiments. We compared MIPTS with Kimi-k2, a state-of-the-art long-text reasoning model, focusing on three key dimensions:

1. Physical Consistency: The ability to recognize physical boundary conditions (e.g., an object will not move backward after coming to a complete stop).
2. Structured Level of Problem-Solving Process: Whether it demonstrates a clear workflow of physical modeling → derivation → result interpretation.
3. Epistemic Rigor: Whether it is equipped with a self-verification mechanism to avoid probabilistic generation hallucinations, and whether its responses feature step-by-step guidance, error correction, and teaching interpretability.

3.1 Experimental Configuration and Description of Case Tasks

To avoid biases arising from a single question type, this study selected six representative university-level physics problems, covering diverse dimensions including kinematics (logical traps), experimental data processing (regression analysis), calculus applications (work done by variable forces), theoretical derivation (inclined planes and circular rings), and conceptual explanation (equivalence principle). All problems were input in natural language without additional prompts or intermediate conclusions, so as to evaluate the system's capabilities of autonomous modeling and reasoning.

3.2 Qualitative Case Studies

To intuitively demonstrate the performance of each model across different physics tasks, we summarize the experimental results in Table 1.

Table 1. Performance Comparison of MIPTS and Baseline Models (Kimi-K2) Across Physics Tasks

Case	Task Domain	Challenge	Baseline Model	MIPTS
1.Braking Problem	Kinematics	Multimodal Perception & Logical Trap	Perception error (v_0 loss); neglect of stop conditions	Visual-semantic alignment + code logic gating
2.Simple Pendulum Experiment	Lab Data	OCR Accuracy & Process Demonstration	"Black-box" output without intermediate steps	Deep document understanding + teaching-level process transparency
3.Work Done by Variable Force	Calculus	Symbolic Integration & Variable Analysis	Correct results but lack of verification	SymPy symbolic computation + manual/code dual verification
4.Inclined Plane Slider	Mechanics	Symbolic Derivation & Limit Checking	Probabilistic text generation	Automated limit analysis ($\theta \to 0, 90^\circ$)
5.Lunar Free Fall	Concepts	Theoretical Depth & Citation	Colloquial expression without formal proof	Structured instruction + RAG knowledge tracing
6.Rotating Circular Ring	High-order Dynamics	Stability Analysis & Complex Reasoning	Prone to reasoning chain breakage	Structured Chain of Thought (CoT) + bifurcation point stability analysis

As shown in Table 1, the various systems exhibit significant discrepancies in their overall performance on the six test questions.

General-purpose model systems can, in most cases, provide formally valid problem-solving steps, yet their reasoning processes are often plagued by the following flaws:

1. Implicit physical assumptions are not explicitly stated;
2. Logical gaps exist in intermediate reasoning steps;
3. There is a lack of self-correction mechanisms when derivation errors occur.

In contrast, the proposed system demonstrates stronger physical consistency and pedagogical structure during the problem-solving process. Its responses typically distinguish clearly between the following phases: Physics scenario analysis → Mathematical model construction → Key formula derivation → Result interpretation and physical significance elaboration.

This structure is highly consistent with the "standard problem-solving paradigm" commonly adopted in university physics teaching, which is more conducive to helping students comprehend physical concepts rather than merely obtaining the final answers.

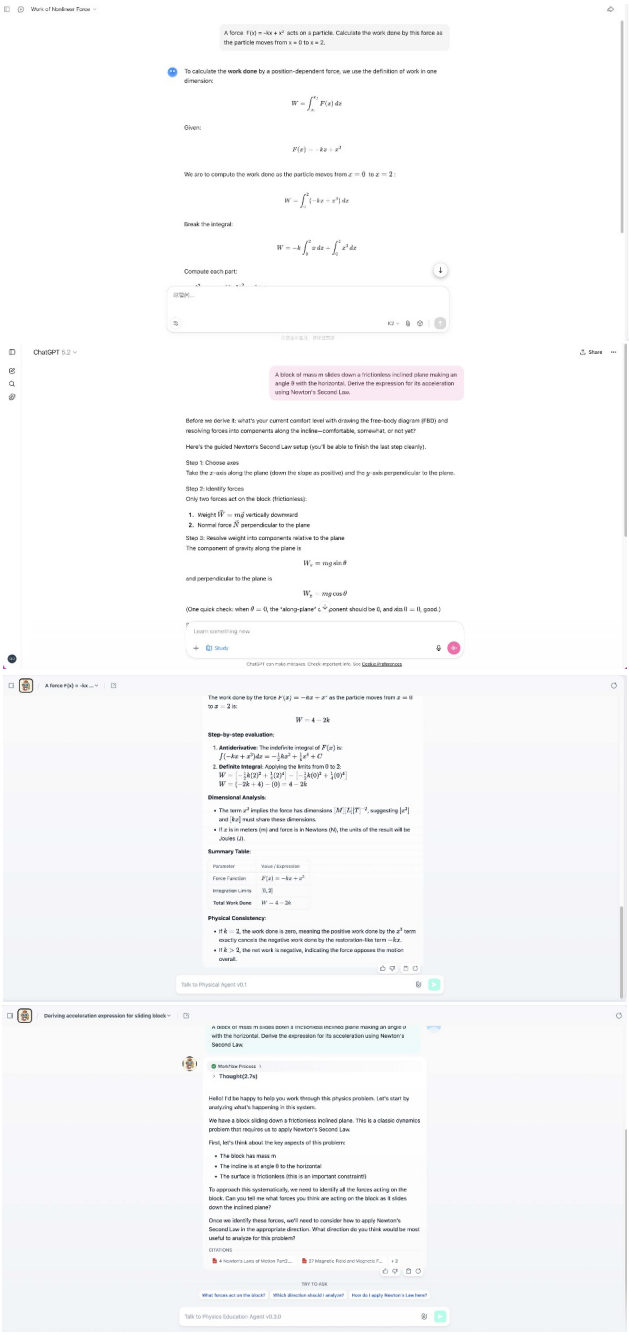

Fig. 3. Multi-Model Interface Overview: A Comparison of Response Styles of Kimi-k2, ChatGPT and MIPTS in an Identical Teaching Scenario.

As shown in Figure 3 is the comparative process diagram. The top two diagrams respectively depict the learning modes driven by the Kimi-k2 0905 Preview model and the GPT-5.2 model. The bottom-left diagram presents the direct answer output of MIPTS, while the bottom-right one displays the continuous question-and-answer interface of the Socratic teaching mode. We selected two of these cases to illustrate the features of MIPTS.

Case 3 & 4: Symbolic Computation and Dual Verification (Variable Force Work and Inclined Plane Derivation) Problem Description: Solve separately for the work done by the variable force $F(x) = -kx + x^3$ (via calculus) and the acceleration of a slider on a frictionless inclined plane (via symbolic derivation).

1. Baseline Model Performance: For the variable force work problem, the baseline model was able to perform integral calculations but lacked discussion on the physical meaning of the symbolic constant k. For the inclined plane problem, the model generated correct textual derivations; however, these were essentially probability-based text completions without mathematically deterministic verification.
2. MIPTS Performance: 1. Symbolic Calculus: MIPTS invokes Python's sympy library for symbolic integration, ensuring absolute precision in mathematical operations. 2. Dual Verification: While outputting the code execution results, the Agent automatically generates manual derivation steps (Manual Verification) and conducts a physical discussion on the result $W = 4 - 2k$ (the impact of the k value on the positive/negative nature of work). 3. Limit Check: For the inclined plane problem, the system automatically performs limit analysis ($a \to 0$ as $\theta \to 0$), emulating the thinking mode of physicists.
3. Analysis: By introducing a symbolic computation engine, MIPTS elevates physical derivation from "probabilistic simulation" to the level of "deterministic computation".

Case 6: Advanced Dynamic Analysis (Small Bead on a Rotating Ring) Problem Description: Analyze the equilibrium point stability of a constrained small bead on a rotating ring (via Lagrangian mechanics / non-inertial frame analysis).

1. Baseline Model Performance: The model was able to formulate the Lagrangian and solve the corresponding equations; however, in the section of stability analysis, its discussion on the bifurcation point ω_c was rather cursory.
2. MIPTS Performance: The system adopted a structured chain-of-thought, explicitly dividing the solution into four steps: force analysis, equilibrium condition derivation, critical angular velocity deduction, and stability criterion formulation. Particularly in stability analysis, it clearly derived the sign variation of the second derivative of the effective potential energy and accurately identified the instability phenomenon of the bottom equilibrium point when $\omega > \omega_c$.
3. Analysis: This case demonstrates that for sophisticated university-level physics problems, MIPTS's structured reasoning framework can organize the problem-solving path more effectively and avoid logical omissions in long-chain reasoning processes.

4 Conclusion and Future Work

We presented MIPTS, a multimodal physics tutoring system that integrates hybrid retrieval (graph-based and deep document understanding) with a tool-augmented reasoning agent under Dify orchestration. The design addresses hallucination, calculation errors, and multimodal blindness by separating perception, routing, retrieval, and verifiable reasoning.

Future work includes larger-scale evaluations with classroom deployment, improved symbolic solvers for advanced mathematics, and extension to video-based physics experiments with multimodal retrieval.

Acknowledgments

This research was funded by the Hubei Provincial Educational Science Planning Project (Project No. 2021ZB121) and the Wuhan College of Arts and Sciences Institutional Teaching Reform Research Project (Project No. 2025JG05).

References

1. Chu, Z., Wang, S., Xie, J., Zhu, T., Yan, Y., Ye, J., Zhong, A., Hu, X., Liang, J., Yu, P.S., & Wen, Q. (2025). LLM Agents for Education: Advances and Applications. ArXiv, abs/2503.11733.
2. Zhang, J. (2025). AI Agents in Education: Four Trends and a Practical Workflow. AAAI Spring Symposia.
3. Shafiq, M., Sami, M.A., Bano, N., Bano, R., & Rashid, M. (2025). Artificial Intelligence in Physics Education: Transforming Learning from Primary to University Level. Indus Journal of Social Sciences.
4. Robledo-Rella, V., & Toh, B.-Y. (2024). Artificial intelligence in physics courses to support active learning. In ICSLT '24: Proceedings of the 2024 10th International Conference on e-Society, e-Learning and e-Technologies (ICSLT) (pp. 68-75). (ICSLT Proceedings). Association for Computing Machinery. https://doi.org/10.1145/3678610.3678631
5. Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng, X., Qin, B., & Liu, T. (2023). A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions. ACM Transactions on Information Systems, 43, 1 - 55.
6. Simon Scheider, Harm M. Bartholomeus, and Judith A. Verstegen, 2023. ChatGPT is not a pocket calculator - Problems of AI-chatbots for teaching Geography, https: //arxiv.org/abs/2307.03196. ICSLT 2024, June 21–23, 2024, Rome, Italy Víctor Robledo-Rella and Bee-Yen Toh
7. Beale, R. (2025). Dialogic Pedagogy for Large Language Models: Aligning Conversational AI with Proven Theories of Learning. ArXiv, abs/2506.19484.
8. Sun, C., Han, J., Deng, W., Wang, X., Qin, Z., & Gould, S. (2023). 3D-GPT: Procedural 3D Modeling with Large Language Models. 2025 International Conference on 3D Vision (3DV), 1253-1263.
9. Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng, X., Qin, B., & Liu, T. (2023). A Survey on Hallucination in Large Language Models: Principles,

Taxonomy, Challenges, and Open Questions. ACM Transactions on Information Systems, 43, 1 - 55.
10. Guo, Z., Xia, L., Yu, Y., Ao, T., & Huang, C. (2024). LightRAG: Simple and Fast Retrieval-Augmented Generation. ArXiv, abs/2410.05779.
11. InfiniFlow. (n.d.). RAGFlow [Computer software]. GitHub. https://github.com/infiniflow/ragflow
12. Fang, J., Peng, Y., Zhang, X., Wang, Y., Yi, X., Zhang, G., Xu, Y., Wu, B., Liu, S., Li, Z., Ren, Z., Aletras, N., Wang, X., Zhou, H., & Meng, Z. (2025). A comprehensive survey of self-evolving AI agents: A new paradigm bridging foundation models and lifelong agentic systems. arXiv. https://arxiv.org/abs/2508.07407
13. Langgenius. (n.d.). Dify: An LLM app development platform [Computer software]. GitHub. https://github.com/langgenius/dify