MobiClues: Constructivist Learning through Gamified Discovery and Multi-Agent AI Tutoring

***Abstract*—This paper presents the design and prototype of MobiClues, a mobile learning application that couples gamified artifact discovery with an intelligent tutoring system (ITS), organizing generative AI and large language models (LLMs) around constructivist learning theory. A learning session alternates between two phases. In the *discovery phase*, learners photograph real-world objects that a vision LLM turns into described virtual items, fuse item pairs via a reasoning LLM and a text-to-image diffusion model, and receive multi-faceted relevance scores as incremental clues toward a hidden artifact of educational significance—anchoring learning in the learner's authentic surroundings using only text and 2D images on a device every student carries. In the *tutoring phase*, a multi-agent retrieval-augmented generation (RAG) ITS diagnoses prior knowledge through LLM-generated probing questions, scaffolds within the learner's zone of proximal development, gates retrieval by pedagogical need, and supports multimodal chat over the learner's own creations, while opt-in peer-collaboration and reflection agents enact social negotiation of meaning and metacognition. Each round feeds the next: the diagnosed knowledge level is re-assessed from dialogue and discovery actions, so scaffolding deepens or fades as competence develops, and knowledge is constructed across iterations of making, feedback, and articulation. We describe the theoretical framework, the system architecture, a prototype walkthrough, and the planned evaluation combining usability and technology-acceptance instruments with interaction-trace analysis for this ongoing project.**

***Index Terms*—large language models, retrieval-augmented generation, constructivist learning, intelligent tutoring systems**

I. Introduction

Constructivist learning theory holds that knowledge is actively constructed rather than transmitted, through exploration, social interaction, and reflection [1], [2]. LLMs offer unprecedented means to operationalize this view at scale [3], yet purely generative tutors hallucinate and rarely adapt to a learner's evolving proficiency, and many AI-enhanced learning environments couple generative AI to heavyweight immersive platforms whose cost can overshadow the pedagogy [4].

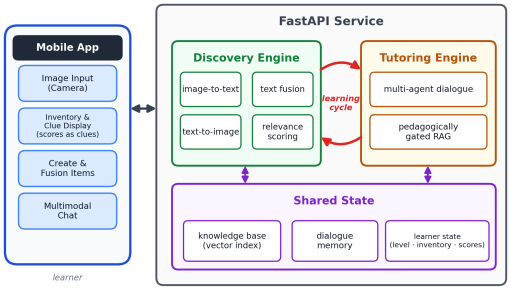


Fig. 1. High-level architecture: a mobile app backed by a FastAPI service hosting a discovery engine and a tutoring engine over shared state; the red cycle marks the learning cycle alternating between them within a session.

MobiClues is designed from the opposite direction: the technology does not dictate learning theory, but rather serves as a means to implement it. A learning session is one iterative cycle with two alternating phases (Fig. 1). In the *discovery phase*, learners photograph real-world objects, which a multimodal LLM transforms into described virtual items; they fuse item pairs, which a reasoning LLM integrates into new descriptions rendered by a text-to-image diffusion model; and each creation receives a multi-faceted relevance score acting as an incremental clue toward a hidden artifact. In the *tutoring phase*, a multi-agent retrieval-augmented generation (RAG) intelligent tutoring system converts those actions into understanding: agents diagnose prior knowledge, scaffold explanations, gate retrieval, and prompt collaboration and reflection—and the learner can attach any created item in chat for the tutor to examine. The phases need each other: creation without dialogue is activity without articulated understanding; dialogue without creation is verbal knowledge that never touches the learner's world. Restricting the interface and interaction on a smartphone is a deliberate theoretical commitment: the mobile camera embeds learning in the learner's authentic surroundings—Honebein's authenticity goal [5]—rather than a simulated environment onscreen.

This paper contributes: (1) **a novel integration of generative AI and RAG for constructivist learning**, in which generative pipelines externalize learner hypotheses as artifacts while retrieval-augmented dialogue grounds their interpretation, every component accountable to a named constructivist principle; (2) **a pedagogically aligned multi-agent intelligent tutoring system** whose cooperating LLM agents map to distinct pedagogical functions, with multimodal chat over learner-created items; and (3) **an integrated multi-round learning session** in which discovery and tutoring alternate under per-round knowledge re-diagnosis. The mobile prototype is implemented and yields preliminary feasibility findings (Section V-A); the user study is planned (Section VI).

II. Related Work

Gamified constructivist learning and generative AI. Gamification supports motivation and engagement [6] but often privileges quantifiable metrics over knowledge construction, a tension visible in gamified constructivist deployments [7]. Generative AI broadens the creative side: AI-driven content fosters collaborative creativity and learning-by-doing [4],

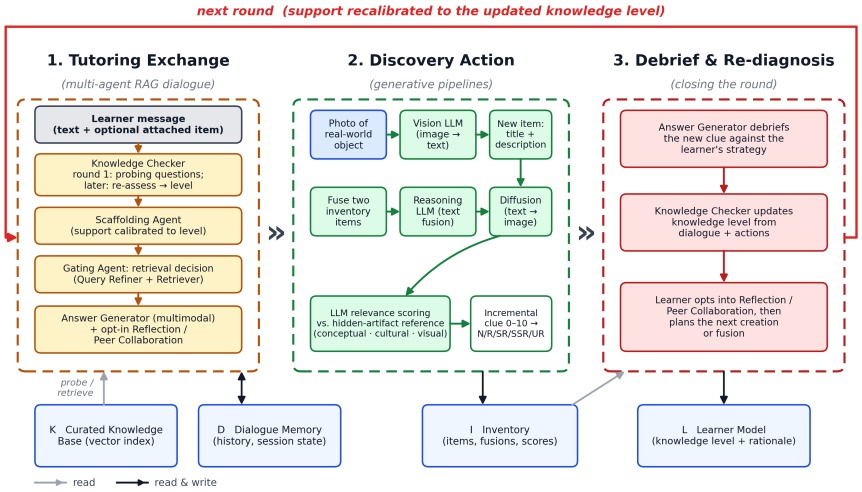


Fig. 2. **Overview of the MobiClues learning cycle.** (**Tutoring Exchange**) The Knowledge Checker diagnoses the learner via probing questions from the knowledge base K, the Scaffolding Agent calibrates support to the diagnosed level, the Gating Agent decides per turn whether to ground the reply in K, and the multimodal Answer Generator synthesizes the response. (**Discovery Action**) The learner photographs an object (vision LLM $\rightarrow$ described item) or fuses two items (reasoning LLM, then text-to-image diffusion); the creation is scored against the hidden artifact, yielding an incremental clue written to the inventory I. (**Debrief & Re-diagnosis**) The clue is debriefed, opt-in Reflection and Peer Collaboration prompt articulation, and the Knowledge Checker updates the learner state L before the next round. D: dialogue memory.

and non-prescriptive co-creation preserves learner ownership [8]. MobiClues combines both threads while resisting the metric trap: its relevance score is an interpretive signal rewarding creative alignment and inviting hypothesis revision [1], with generative interaction tied to explicit learning objectives.

Adaptive RAG and multi-agent LLM systems. RAG grounds generation in verifiable sources [9]; adaptive variants condition retrieval on question complexity [10] or let the model decide when to retrieve and self-critique [11], optimizing factual answering with single-agent designs. Multi-agent work shows role specialization improves interpretability and control [12]. MobiClues extends both lines to intelligent tutoring: agents map to pedagogical functions, and gating weighs the learner's diagnosed knowledge state and instructional trajectory—not just query complexity—so retrieval serves scaffolding rather than mere answer accuracy.

III. Pedagogical Framework

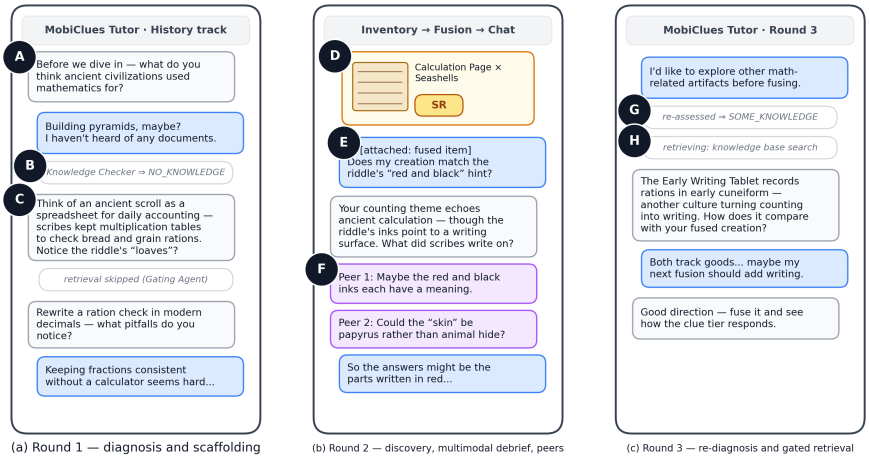
Four constructivist tenets drive every design decision. **Prior-knowledge activation and diagnosis**: learning is most effective when new information links to existing schema and misconceptions surface early [1]; the Knowledge Checker opens each topic with corpus-grounded probing questions, assigns a label (*NO_KNOWLEDGE*, *SOME_KNOWLEDGE*, *KNOWLEDGEABLE*) that conditions all downstream agents, and re-assesses it every round. **Scaffolding within the zone of proximal development**: a dedicated Scaffolding Agent calibrates the depth, complexity, and language of every explanation to the diagnosed state, fading as competence emerges across rounds. **Social negotiation of meaning**: understanding is co-constructed through dialogue and alternative perspectives [2]; learners exchange created items, and an opt-in Peer Collaboration agent simulates peer voices that pose questions, surface misconceptions, and offer competing readings. **Metacognition and reflection**: articulating and evaluating one's own strategy converts activity into self-regulated knowledge; the opt-in Reflection Agent prompts self-explanation at meaningful junctures, and each clue score invites the learner to ask *why* their hypothesis moved toward or away from the target.

IV. System Architecture: An Integrated Learning Session

MobiClues is a mobile application backed by a FastAPI service (Fig. 1). The app provides the learner-facing functions: *item creation*, turning camera photos into described virtual items; *item fusion* over a personal inventory; *multimodal chat*; and *clue display*. The service hosts a *discovery engine* (generative pipelines) and a *tutoring engine* (multi-agent dialogue, orchestrated with `langgraph`) over shared state—curated knowledge base (vector index), dialogue memory, learner state, and inventory. The engines implement the two phases of one learning cycle (Fig. 2), coupled through the shared state each reads and writes every round; Fig. 3 shows these functions in use.

A. One Cycle, Two Phases

The role of the discovery phase. Constructivist theory demands that learners *act*, yet in most LLM-based learning tools the only action is asking another question. The generative pipelines supply a richer action vocabulary: photographing an object is an interpretive act made explicit by the generated title and description, and fusing two items lets the diffusion model render the learner's conceptual combination back as a percept—the learner *sees* their hypothesis, and the relevance score converts it into calibrated feedback. **The role of the**



(a) Round 1 — diagnosis and scaffolding (b) Round 2 — discovery, multimodal debrief, peers (c) Round 3 — re-diagnosis and gated retrieval

Fig. 3. A three-round MobiClues session in the mobile chat interface; circled letters (A)–(H) mark internal agent decisions, and the italic status labels visualize these internal traces for illustration only—they are not part of the learner-facing interface. (a) **Round 1**: probing questions (A), NO_KNOWLEDGE assessment (B), foundational scaffolding with retrieval skipped (C). (b) **Round 2**: fusion scored **SR** (D), learner attaches the creation for multimodal critique (E), simulated peer voices (F). (c) **Round 3**: re-diagnosis to SOME_KNOWLEDGE (G), gated retrieval grounding the reply in the *Early Writing Tablet* (H).

tutoring phase. A score of SR tells the learner *that* a direction is productive, not *why*; the tutoring phase supplies the interpretive layer—debriefing each clue, connecting creations to domain concepts, grounding claims in retrieved sources when warranted. **The cycle.** A session is a sequence of rounds: a tutoring exchange (opening with diagnosis in round one), a discovery action whose score becomes the next incremental clue, and a debrief with optional reflection or peer collaboration; the Knowledge Checker then updates the learner's label, and rounds repeat until a final item is submitted and the hidden artifact is revealed with a closing reflection.

B. Discovery Phase: Generative Item Creation and Fusion

Learners photograph a real-world object; a multimodal LLM (`gpt-5.6-luna`) generates a concise title and text description, and photograph plus text enter the inventory as a new *item*—real-world contextualization without simulation. To pursue the hidden artifact, learners select two items: the same model integrates their attributes into a fused title and description, visualized by a text-to-image model (`gpt-image-2`); the fused item re-enters the inventory for further fusion. Every new or fused item is scored by a stronger reasoning tier (`gpt-5.6-terra`) against the hidden artifact's reference content along conceptual, cultural, and visual-descriptive dimensions, yielding a 0–10 score mapped to rarity tiers (**N/R/SR/SSR/UR**). Scores function as *incremental clues*—a rising score signals a productive direction, a falling one invites revision—and since high scores can arise from replication *or* imaginative conceptual alignment, the metric legitimizes multiple perspectives; each clue passes immediately to the tutoring phase, so assessment opens the next round.

C. Tutoring Phase: Multi-Agent Retrieval-Augmented Dialogue

All dialogue flows through the multi-agent pipeline of Fig. 2; each agent is a dedicated system prompt conditioning `gpt-5.6-luna`, whose multimodal input supports attached-item chat and whose prompt caching keeps per-turn agent calls inexpensive; pedagogical strategies are tuned by editing prompts. The **Knowledge Checker** retrieves representative corpus passages on entering a topic and generates two to three target-knowledge probing questions spanning recall, conceptual, and applied levels; it assesses the learner's replies against the reference content and assigns a label with a stored rationale, doubling as prior-knowledge activation. In later rounds it updates the label from evolving evidence—question sophistication, fusion coherence, attached items—without re-quizzing. This training-free design is an implementation decision—a new domain requires only a curated corpus—whose consistency the planned study will examine. The **Scaffolding Agent** calibrates hints, analogies, and depth to the diagnosed state; the **Gating Agent** decides per turn whether retrieval is warranted [10], [11]—skipping it for conversational turns, triggering it for factual queries—after which the **Query Refiner** and **Retriever** fetch passages by vector similarity. The **Answer Generator** synthesizes each reply from scaffolding directives, retrieved evidence, any *attached item's image and description*, and dialogue history, embedding metacognitive prompts; the opt-in **Peer Collaboration** and **Reflection** agents enact social negotiation and self-explanation. Because session state includes the inventory and score history, the tutoring phase is always aware of the discovery phase; multimodal chat thus makes the learner's creations first-class citizens of the dialogue.

V. Prototype Walkthrough

We illustrate a session in the *History* track, whose corpus is drawn from *A History of the World in 100 Objects* [13]. The hidden artifact is the **Rhind Mathematical Papyrus**—an Egyptian scroll of eighty-four arithmetic problems, questions in black, answers in red; the learner never sees it and receives only a riddle:

> **In black the questions, in red where answers begin;**
> **loaves, grain, and fractions—every sum is mine.**

Fig. 3 shows a three-round excerpt in the mobile chat interface, with circled letters (A)–(H) marking internal agent decisions. In Round 1, probing questions (A) yield a NO_KNOWLEDGE assessment (B), prompting foundational scaffolding via the spreadsheet analogy (C) with retrieval skipped. In Round 2, the learner fuses a photographed calculation page with a *Seashells* item, scored **SR** (D), attaches the creation in chat for multimodal critique against the riddle (E), and debates it with simulated peers (F). In Round 3, re-diagnosis promotes the learner to SOME_KNOWLEDGE (G) and gating triggers retrieval (H), grounding the reply in the *Early Writing Tablet*. Diagnosis anchors prior knowledge, the fusion score acts as an incremental clue, the attached-item and peer exchanges enact social negotiation over the learner's actual creation, and retrieval occurs exactly once—when grounding becomes pedagogically necessary.

A. Preliminary Findings

The prototype yields three preliminary findings. **(1) Feasibility.** The full constructivist cycle runs end-to-end on a commodity smartphone against hosted model APIs alone: no fine-tuning, training data, or dedicated hardware; a new domain reduces to a curated corpus plus prompt edits. **(2) Cost–capability split.** A low-cost model tier suffices for all high-volume calls; a stronger tier is warranted only for relevance scoring; with prompt caching over shared agent context, the multi-agent decomposition stays inexpensive per turn and appears practical at classroom scale. **(3) Auditability.** Every pedagogical decision is an explicit agent output—labels with rationales, gating decisions, dimension-wise scores—so sessions are inspectable end-to-end (visualized as the status labels of Fig. 3 in internal-testing builds), which aided development and enables the trace analysis of Section VI.

VI. Evaluation Design

We plan a user study with university students combining a pre-questionnaire, a post-questionnaire based on the System Usability Scale [14] and the Technology Acceptance Model [15], and a framework-specific questionnaire on four pillars: *Authentic Mobile Contextualization*, *Active Exploration*, *Artifact Discovery*, and *Adaptive Tutoring Dialogue*. Logged traces—knowledge-label trajectories, item attachments, gating decisions, score progressions—will examine whether LLM-based diagnosis matches expert judgments, whether scaffolding calibrates to and fades with the diagnosed state, whether gating reduces unnecessary retrieval, and whether score progressions correlate with self-reported gains; a reveal session hosts the closing reflection and post-questionnaires.

VII. Conclusion

MobiClues integrates a generative-AI discovery phase and a multi-agent retrieval-augmented tutoring phase into one iterative learning session under an explicit constructivist design discipline: probing questions diagnose prior knowledge, scaffolding operates within the zone of proximal development, peers and item exchange enact social negotiation, reflection is embedded in discovery, and multimodal chat makes the learner's creations first-class citizens of the dialogue. Beyond the upcoming user study, limitations include validating LLM-based diagnosis and calibrating LLM-judged scoring against educators; future work envisions persona-based tutoring styles and peer item exchange. MobiClues offers a blueprint for generative-AI learning environments in which learning theory, rather than technological spectacle, drives the architecture.

References

[1] S. O. Bada, "Constructivism Learning Theory: A Paradigm for Teaching and Learning," *IOSR Journal of Research & Method in Education*, vol. 5, no. 6, pp. 66–70, 2015.

[2] M. Tam, "Constructivism, instructional design, and technology: Implications for transforming distance learning," *Educational Technology & Society*, vol. 3, no. 2, pp. 50–60, 2000.

[3] E. Kasneci, K. Seßler, S. Küchemann, M. Bannert, et al., "ChatGPT for good? On opportunities and challenges of large language models for education," *Learning and Individual Differences*, vol. 103, p. 102274, 2023.

[4] G. Vallasciani, L. Stacchio, P. Cascarano, and G. Marfia, "CreAIXR: Fostering Creativity with Generative AI in XR environments," in *Proc. IEEE MetaCom*, 2024, pp. 1–8.

[5] P. C. Honebein, "Seven goals for the design of constructivist learning environments," in *Constructivist Learning Environments: Case Studies in Instructional Design*, B. G. Wilson, Ed. Englewood Cliffs, NJ: Educational Technology Publications, 1996, pp. 11–24.

[6] M. Sailer and L. Homner, "The Gamification of Learning: a Meta-analysis," *Educational Psychology Review*, vol. 32, no. 1, pp. 77–112, 2020.

[7] P. H. F. Ng, F. T. K. Har, K. S. K. Tai, W. C. L. Leung, et al., "Gamified Constructivist Teaching in Metaverse: Revolutionizing Language Learning in University via an Immersive Experience," in *Proc. IEEE MetaCom*, 2024, pp. 101–108.

[8] K. I. Gero, T. Long, and L. B. Chilton, "Social Dynamics of AI Support in Creative Writing," in *Proc. ACM CHI*, 2023, pp. 1–15.

[9] P. Lewis, E. Perez, A. Piktus, F. Petroni, et al., "Retrieval-augmented generation for knowledge-intensive NLP tasks," in *Proc. NeurIPS*, vol. 33, 2020, pp. 9459–9474.

[10] S. Jeong, J. Baek, S. Cho, S. J. Hwang, and J. C. Park, "Adaptive-RAG: Learning to Adapt Retrieval-Augmented Large Language Models through Question Complexity," in *Proc. NAACL*, 2024.

[11] A. Asai, Z. Wu, Y. Wang, A. Sil, and H. Hajishirzi, "Self-RAG: Learning to Retrieve, Generate, and Critique through Self-Reflection," in *Proc. ICLR*, 2024.

[12] J. S. Park, J. C. O'Brien, C. J. Cai, M. R. Morris, P. Liang, and M. S. Bernstein, "Generative Agents: Interactive Simulacra of Human Behavior," in *Proc. ACM UIST*, 2023, pp. 1–22.

[13] N. MacGregor, *A History of the World in 100 Objects*. New York, NY, USA: Penguin Books, 2013.

[14] P. Vlachogianni and N. Tselios, "Perceived usability evaluation of educational technology using the system usability scale (SUS): A systematic review," *Journal of Research on Technology in Education*, vol. 54, no. 3, pp. 392–409, 2022.

[15] V. Venkatesh and H. Bala, "Technology acceptance model 3 and a research agenda on interventions," *Decision Sciences*, vol. 39, no. 2, pp. 273–315, 2008.