

Bridging the Learning Divide with StudyAI

Caroline Rennie, University of Edinburgh, s2878824@ed.ac.uk

Orkun Kınay, University of Edinburgh, s2892215@ed.ac.uk

[Demo link](#)

Abstract

High-quality tutoring improves student performance significantly but remains accessible to only 2% of U.S. students due to cost barriers. Current AI tutors fail to replicate effective human teaching, providing reactive answers that foster "hallucinated understanding" rather than genuine mastery. StudyAI addresses this by applying Agentic Context Engineering to education: integrating course materials into an evolving knowledge base that expands rather than compresses understanding. Through Socratic dialogue, personalized adaptation, and multi-model verification, the system provides course-specific support while maintaining academic integrity. Designed for open-source institutional deployment, StudyAI transforms tutoring from an expensive privilege into a scalable educational resource accessible to all students.

Introduction / Problem Statement

High-quality tutoring is transformative yet fundamentally inequitable. Research demonstrates that personalized tutoring adds 3-15 months of learning per year, improving student performance by 0.37 standard deviations, sufficient to move a typical student from the 50th to the 64th percentile [4]. However, a 2023 USC Schaeffer Center study found that only 2% of U.S. students receive quality tutoring, with access overwhelmingly concentrated among wealthy families who can afford £40-100 per hour [3]. This creates a reinforcing cycle where financial privilege, rather than academic potential, determines educational outcomes.

While generative AI promises "24/7 tutoring," current implementations fail to replicate effective human teaching. Existing AI tutors provide reactive answers disconnected from course materials, requiring constant prompt engineering and piecemeal content uploads. They cannot maintain coherent understanding across lectures or diagnose true comprehension, resulting in "hallucinated understanding", students feel confident without genuine mastery.

A 2024 Harvard study revealed the path forward: when researchers carefully tailored an AI tutor to a physics course using actual course materials, student engagement and learning gains doubled compared to traditional active-learning classrooms. However, the study emphasized that such outcomes require thoughtful architectural design [2]. The underlying technical challenge was identified by Stanford researchers in their 2025 paper on Agentic Context Engineering (ACE): AI systems suffer from *brevity bias*, repeatedly condensing information until critical pedagogical scaffolding disappears. In their experiments, this context collapse reduced reasoning performance from 66.7% to 57.1% in a single iteration. Current AI tutors summarize when they should teach, erasing the depth required for understanding.

Proposed Solution

StudyAI applies Agentic Context Engineering to education to overcome these limitations. Rather than compressing course knowledge, our system expands and refines understanding through continuous interaction. Students upload their syllabus, lecture slides, and course materials once; StudyAI integrates them into a unified, evolving knowledge base that grows more personalized with each session.

Personalized Adaptation: StudyAI diagnoses baseline understanding through conversational questioning, establishing confidence levels and pacing preferences. All explanations draw from the student's actual course materials, matching instructor notation exactly. The system builds a "study playbook" capturing misconceptions and effective strategies, refining support while preserving pedagogical detail.

Interactive Socratic Learning: Rather than delivering definitions, StudyAI engages students through inquiry-driven dialogue [1]. Questions like "Imagine testing for a rare disease with a 99% accurate test, your result is positive. What's the chance you actually have it?" encourage reasoning and reflection, anchoring abstract concepts to familiar domains.

Integrated Course Context: Unlike traditional AI tutors that treat queries as isolated events, StudyAI unifies lecture slides, readings, and assignments into a single evolving context. Explanations reference specific materials, maintaining notation consistency and enabling smooth conceptual transitions across the curriculum.

Academic Integrity by Design: When students upload assignments, the system responds: "I'll help you build the concepts, but won't solve these exact problems." Practice problems test understanding through varied scenarios, with reflection prompts and internal verification to reduce overconfident errors.

StudyAI operates on a retrieval-augmented generation (RAG) architecture with multi-model verification. Course materials are processed through a pipeline that chunks content semantically, embeds, and indexes for sub-second context retrieval. When students ask questions, the system retrieves relevant sections and generates explanations, then a model from another provider performs independent fact-checking and depth verification [5]. Disagreements between models trigger additional retrieval and synthesis, with confidence scores included in every response for transparency.

The system comprises three agentic modules: a Generator (creates explanations grounded in retrieved context), a Reflector (validates accuracy and pedagogical depth), and a Curator (maintains the evolving knowledge base) [6]. This architecture prevents the context collapse that plagues simpler systems, instead of summarizing away detail, StudyAI accumulates richer understanding over time, mirroring how expert human tutors refine their teaching.

Our current prototype uses commercial APIs (OpenAI and Anthropic) for rapid development and validation. However, for institutional-scale deployment, we plan to transition to smaller fine-tuned open-source models such as Llama 4 or Qwen 3, which would dramatically reduce operational costs while maintaining control over model behavior and ensuring student data never leaves institutional infrastructure. Because our serving layer is already standardized on LiteLLM, switching is operationally trivial. This approach would enable on-premise deployment, addressing data sovereignty concerns and reducing per-student costs from approximately £5 per tutoring hour to under £1 through self-hosted inference on university GPU clusters.

Implications & Ethical Considerations

We acknowledge critical risks in AI tutoring and have designed explicit mitigations. Over-reliance is addressed through reflective prompts that ask students to summarize, transfer concepts to new contexts, and self-check rather than simply obtaining answers. Drift and forgetfulness (context collapse) are curbed by incremental, auditable updates to the learner's playbook that preserve pedagogical detail rather than compressing it. Incorrect confidence is reduced through internal challenge steps where the system critiques its own explanations before presenting them [5]. Bias and fairness concerns are mitigated by grounding explanations in student-provided course materials rather than relying solely on model priors that may reflect majority-culture assumptions. Equity of access is further supported through phone-friendly interfaces and low-bandwidth compatibility, ensuring that students who lack regular one-to-one help can still receive comparable educational benefits.

References

- [1] Hu, X., Xu, S., Tong, R., & Graesser, A. (2025). *Generative AI in education: From foundational insights to the Socratic playground for learning* (arXiv:2501.06682). arXiv. <https://arxiv.org/abs/2501.06682>
- [2] Manning, A. J. (2024, September 5). *Professor tailored AI tutor to physics course: Engagement doubled*. *Harvard Gazette*. <https://news.harvard.edu/gazette/story/2024/09/professor-tailored-ai-tutor-to-physics-course-engagement-doubled/>
- [3] Rapaport, A., & Silver, D. (2023, March 10). *Two percent of U.S. children receive high quality tutoring, despite billions funneled into school systems*. *USC Schaeffer Center for Health Policy & Economics*. <https://schaeffer.usc.edu/research/two-percent-of-u-s-children-receive-high-quality-tutoring-despite-billions-funneled-into-school-systems/>
- [4] Shoemaker DeMio, P. (2024, January 18). *Scaling up high-dosage tutoring is crucial to students' academic success*. *Center for American Progress*. <https://www.americanprogress.org/article/scaling-up-high-dosage-tutoring-is-crucial-to-students-academic-success/>
- [5] Xu, X., Chen, T., Zhang, F., Liu, W., Li, P., Jaiswal, A. K., Yan, Y., Hu, J., Wang, Y., Chen, H., & others. (2025). *Double-checker: Enhancing reasoning of slow-thinking LLMs via self-critical fine-tuning* (arXiv:2506.21285). arXiv. <https://arxiv.org/abs/2506.21285>
- [6] Zhang, Q., Hu, C., Upasani, S., Ma, B., Hong, F., Kamanuru, V., Rainton, J., Wu, C., Ji, M., Li, H., & others. (2025). *Agentic context engineering: Evolving contexts for self-improving language models* (arXiv:2510.04618). arXiv. <https://arxiv.org/abs/2510.04618>