

From Surface Learning to Deep Understanding: A Grounded AI Tutoring System for Moodle

Anna Ostrowska , Michał Kukla , Gabriela Majstrak , Jan Opala , Sebastian Pergała , Jan Skwarek , Anna Wróblewska

Faculty of Mathematics and Information Sciences, Warsaw University of Technology, Warsaw, Poland
anna.wroblewska1@pw.edu.pl

Abstract

This demo paper describes the development of the AI Teaching & Learning Tutor Assistant, a modular Moodle plugin that leverages Retrieval-Augmented Generation (RAG) to deliver high-quality, highly selective answers strictly grounded in teachers' materials. The system employs a dual-centric design, providing students with interactive, Socratic-based tutoring options and educators with a "human-in-the-loop" workspace for supervised content generation, e.g., quizzes, tasks, or short summaries. By grounding Large Language Model (LLM) responses in teacher-provided materials, the assistant addresses the risks of misinformation while encouraging deep conceptual mastery. Evaluation via the LLM-as-a-Judge framework and a preliminary user study confirms its effectiveness, achieving faithfulness scores up to 0.97 and a 4.00/5.00 recommendation rate.

Demo Video: <https://tinyurl.com/mrytc2dr>

1 Introduction

Modern e-learning platforms like Moodle have revolutionized access to education, but they have also introduced information overload. Students often face a massive volume of files and deadlines, leading to "surface learning", a strategy where learners focus on meeting immediate requirements rather than achieving a profound understanding of the subject [Motlagh *et al.*, 2023]. Concurrently, while general AI tools have become popular, they present significant risks: they are prone to hallucinations (generating factually incorrect but plausible-sounding information), they lack institutional integration, and high-quality versions are often expensive, creating an accessibility gap [Dan *et al.*, 2025].

The primary objective of this project is to transform AI from a passive information provider into an active, supervised learning partner. The system aims to: **minimize misinformation** by grounding all AI outputs in verified course materials; **reduce teacher workload** through the semi-automation of "tedious" tasks like a quiz and a study guide preparation; foster **critical thinking** by using pedagogical frameworks like the Socratic method and cognitive content assessment, i.e.,

Bloom's Taxonomy [Valiathan, 2024]; ensure **data privacy** through a modular architecture that isolates data between different courses. Our application follows a dual-centric approach: **Teacher-Centric Use Cases:** The system facilitates the didactic process by automating the preparation of study materials and tests. It maintains a "Human-in-the-loop" workflow, where teachers act as essential validators, reviewing and approving AI-generated content before it is released to students. **Student-Centric Use Cases:** For learners, the assistant provides an interactive chatbot and a personalized quiz module. These features are designed to move students away from passive "surface learning" toward a deep understanding of concepts and long-term retention.

Our assistant is designed to be cost-effective and easy to set up. The cost-effectiveness relates to the smaller LLMs used in our chatbot; they work effectively because of our methods, built on designed prompting strategies and RAG, using the source materials database. The assistant employs a modular design with a FastAPI backend, a Next.js frontend, and a PHP-based Moodle plugin, supporting both local deployment and offloading external computation to ensure scalability.

2 Related Work & Market Analysis

The **evolution of artificial intelligence in education** has shifted from early rule-based and pattern-matching systems, such as ELIZA [Pérez *et al.*, 2020], which lacked conversational memory and produced deterministic responses, to more advanced systems that use Latent Semantic Analysis (LSA) to understand student responses through semantic similarity and adapt teaching strategies. Modern educational chatbots have further advanced this field by combining conversational modules with information retrieval components, enabling dynamic updates to knowledge bases with new academic materials [Pérez *et al.*, 2020]. In current research, a primary methodological distinction is made between fine-tuning, which modifies a model's internal weights at a high computational cost, and RAG, which grounds LLMs' responses in external, non-parametric memory. RAG is increasingly preferred in academic contexts because it enables easy updates to course content and significantly mitigates the risk of hallucinations by providing verifiable citations to teacher-provided sources.

To maintain **high pedagogical standards**, modern systems integrate established strategies such as the Socratic method,

which encourages active thinking by guiding students toward discovery rather than providing direct answers, and Bloom’s Taxonomy, which structures questions across different cognitive levels [Güldal and Dinçer, 2025]. The evaluation of these RAG-based educational tools has also moved beyond traditional Natural Language Processing metrics toward specialized frameworks such as LLM-as-a-Judge, which assess performance across dimensions of faithfulness, answer relevance, and context recall [Dan *et al.*, 2025; Ji *et al.*, 2023].

Currently, **the market** for AI educational tools includes specialized platforms such as StudyFetch [StudyFetch, 2026], NotebookLM, and Quizbot [Quizlet, 2023], which generate learning resources, such as flashcards and quizzes, from uploaded documents. Additionally, major providers offer institutional solutions [Google for Education, 2025; Microsoft Learn, 2025], designed to increase academic productivity and summarize texts.

Despite these advancements, a significant **“universal integration gap”** persists, as most existing prototypes operate as standalone services that require manual file uploads and lack native integration with established Learning Management Systems (LMS) such as Moodle [Hunt, 2012]. Furthermore, there is a critical shortage of open-source systems that support a “human-in-the-loop” workflow, in which teachers act as essential validators, reviewing and approving AI-generated content before it is released to students. The research presented here addresses these shortcomings by offering a modular, cost-effective solution that transforms AI from a passive information provider into an active, teacher-supervised partner in the learning process.

3 Our AI Tutor Assistant Architecture

The system employs a three-layer – **Presentation, Application, and Data** – modular architecture designed for scalability and easy-reproducible deployment.

Presentation Layer (Moodle plugin). The primary interface is built around a native PHP-based Moodle Activity Module plugin [Moodle Project, 2025], which serves as the secure entry point for all users. For teachers, this layer serves as an administrative workspace with a “human-in-the-loop” editor, where AI-generated materials and quizzes are held in “unreviewed” status until they are manually validated and approved. For students, the layer uses a Next.js and React frontend embedded in a Moodle IFrame to provide a responsive environment for interactive chatbots and self-assessment. To ensure academic content remains legible, the presentation layer integrates libraries like KaTeX and Marked.js to accurately render complex LaTeX mathematical formulas and Markdown-formatted text.

Application Layer (FastAPI Backend). A Python-based FastAPI backend serves as the API Gateway, coordinating the RAG pipeline and AI services. This layer handles: **parsing** – extracting text from PDF, DOCX, and PPTX files; **intent recognition** – a lightweight all-MiniLM-L6-v2 sentence-transformer model automatically classifies user prompts into “Explanation,” “Test Generation,” or “Material Generation” modes; **asynchronicity** – all pipelines run asynchronously to

prevent users from blocking one another during long-running tasks like document indexing.

Data Layer (Hybrid Storage). To manage diverse data types and roles, a hybrid strategy is used: **PostgreSQL** stores structured relational data (course info, user metadata, conversation logs); **ChromaDB** – an AI-native vector database that stores high-dimensional embeddings of the source teaching materials for rapid semantic search; **MinIO** – an S3-compatible blob storage for large raw files, keeping the relational database performant.

Deployment and Orchestration. The entire integrated environment is containerized using Docker, allowing the various services –including Moodle, FastAPI, the databases, and storage servers – to run in isolated, reproducible environments.

4 Interaction Design & Interface

The interaction design and interface of the AI Assistant provide a seamless, integrated experience within the **Moodle Learning Management System (LMS)** that supports both teaching and learning workflows, enabling the user to assess the generated content critically at every step (see also [Wróblewska *et al.*, 2025]).

Student-Centric Interface. Students access the system via a home page hub featuring three workflows. **Conversational Chatbot** features three modes: Quick Answer (fact-based), Deep Understanding (Socratic guidance), and Exam Coach (Chain-of-Thought reasoning). Additionally, **Source Reader** allows students to expand AI responses to see exact page numbers and document fragments from within the grounded teacher’s materials. **Quiz Center** generates personalized tests based on the entire course or specific topics. After submission, students receive detailed explanations linked to course materials. **Learning Dashboard** visualizes progress through non-linear coverage metrics, showing which materials are “Completed” or “In Progress” based on interaction frequency.

Teacher-Centric Workflow. Educators use the AI Assistant tab to manage course context. **Human-in-the-loop Editing:** ensures that all AI-generated tests are initially “Unreviewed”. Teachers use a built-in markdown editor to refine questions. **Moodle Integration** enables validated assessments to be exported in the Moodle XML format, allowing them to be imported directly into the institutional Question Bank. To further personalize the environment, a settings panel lets users adjust the visual atmosphere ensuring the tool remains accessible and engaging across devices.

5 Experimental Evaluation

Technical Optimization (LLM-as-a-Judge). The evaluation of the system using the RAGAS (Retrieval-Augmented Generation Assessment) framework provided quantitative insights into the performance of the retrieval and generation pipelines across different academic disciplines and interaction modes [Es *et al.*, 2024].

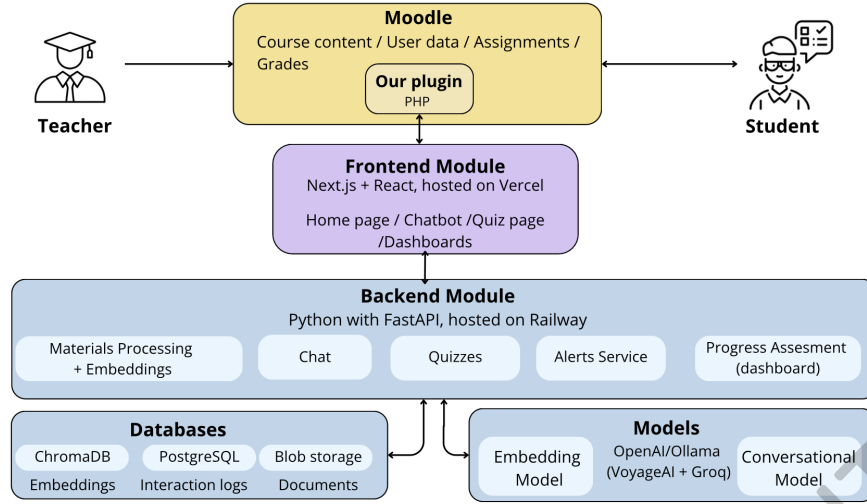


Figure 1: Our system architecture: Presentation Layer – Moodle Plugin and Frontend Module, Application Layer – Backend Module, Data Layer – Databases and Models

Our experiments¹ revealed that the optimal system configuration differs significantly depending on the subject matter: for STEM subjects: The optimal chunk size was 512 characters, and the best performance was reached with a model temperature of 0.3 were optimal for technical formulas, achieving a RAGAS faithfulness of 0.97; for humanities: larger chunks (1,000 chars) and a lower temperature (0.1) were required to capture interpreted meanings, achieving a context recall of 1.00, and an average faithfulness of 0.98. Across both disciplines, Gemini Flash 2.0 was chosen for production due to its superior faithfulness and speed-to-cost ratio compared to GPT-4o-mini and Llama 3.1. For retrieval settings, the “sweet spot” for context was found to be Top-K = 10; increasing this to 15 did not improve recall but significantly worsened context precision due to the “distraction issue” caused by irrelevant information.

The Ragas evaluation of the three specialized interaction modes confirmed that they successfully achieved pedagogical goals: **Quick Mode** achieved the highest faithfulness (0.92), prioritizing direct, factual answers grounded strictly in the source materials; **Deep Understanding** intentionally recorded low answer relevancy because its Socratic design focuses on asking guided questions rather than providing answers, yet it maintained near-perfect context recall (0.98); **Exam Coach** successfully balanced structured support with accuracy, yielding the highest answer relevancy (0.82) through the use of Chain-of-Thought reasoning.

Human Feedback Analysis. The experiments with human feedback evaluated our AI Assistant’s usability, effectiveness, and pedagogical value from both students’ and educators’ perspectives.² These studies utilized a Likert scale (1–5) and open-ended qualitative questions to gather detailed insights

¹The detailed reports on our experiments: [Ostrowska *et al.*, 2026; Kukla *et al.*, 2026]

²We have ethical approval from our university’s Ethics Board to conduct experiments with users and to administer the survey.

into the user experience. The evaluation involved 18 participants in the final study, of whom the majority were students (83.3%) and instructors (11.1%) from our university. Participants were tasked with interacting with the chatbot, generating and solving quizzes based on real academic materials (such as NLP and Machine Learning course notes), and navigating the learning dashboard to track their progress. The system was highly successful in its primary goal of providing grounded, reliable information. Participants consistently rated the resistance to hallucinations as high (Mean = 4.44/5), confirming that RAG was effectively limited to teacher-provided sources. The overall recommendation score for the system was 4.00 out of 5, with the highest individual ratings given to the ease of starting a conversation (4.39/5) and the relevance and accuracy of responses (4.06/5). Furthermore, 14 out of 18 users identified the chat interface and automatic test generation as the most useful features of the assistant and the source reader for academic transparency. Teachers valued the ability to monitor student logs and maintain control over content generation.

6 Summary and Impact

Our AI Assistant bridges the gap between conversational AI and institutional Learning Management Systems. By providing an open-source, cost-effective solution, the assistant eliminates financial barriers to high-quality AI education while maintaining strict data isolation and privacy between courses. For educators, the system drastically reduces the time-intensive burden of manual content creation without sacrificing control, as every AI-generated quiz or study guide must be manually approved before students receive it. By strictly grounding AI in teacher-verified documents and a Socratic chatbot, the system moves students away from “surface learning” toward a grounded mastery of their academic subjects. Ultimately, this project transforms AI from a passive information provider into an active, supervised partner in the learning process.

Acknowledgments

This work was supported by the European Union under the Horizon Europe grant "Overcoming Multilevel Information Overload" (OMINO, <https://ominoproject.eu>, grant no. 101086321) and by the Polish Ministry of Education and Science within the framework of the program titled International Projects Co-Financed.³ We also thank Dr. Daniel Dan (Modul University Vienna) for providing an initial introduction to the task of developing software integrated with Moodle LMS, and the survey participants for testing our system and providing valuable feedback.

Contribution Statement

Anna Ostrowska: System architecture and full-stack integration, frontend, Moodle plugin, quiz, and progress assessment modules, designed pedagogical chatbot modes and handled prompt engineering; **Michał Kukla:** Integrated PostgreSQL and vector databases, orchestrated LLM services, frontend for content editing and Moodle XML; **Gabriela Majstrak:** RAG mechanism, data processing, chunking, and reranking, database integration, parameter tuning, Ragas evaluation; **Jan Opala:** Global infrastructure and server deployment, Moodle integration, alerts service, unit testing; **Sebastian Pergała:** File ingestion pipeline, retrieval logic, Moodle environment, user interface; **Jan Skwarek:** System architecture and database schemes, cloud deployment, secure FastAPI backend, full-stack system integration; **Anna Wróblewska:** Scientific lead, supervision, strategy, and funding. All authors have read and agreed to the published version of the manuscript.

References

- [Dan *et al.*, 2025] Daniel Dan, Anna Wróblewska, Bartosz Grabek, Michał Taczała, and Minoru Nakayama. Applications and challenges of artificial intelligence in educational course design and delivery. In *2025 20th Conference on Computer Science and Intelligence Systems (FedCSIS)*, pages 675–680. IEEE, 2025.
- [Es *et al.*, 2024] Shahul Es, Jithin James, Luis Espinosa Anke, and Steven Schockaert. RAGAS: Automated evaluation of retrieval augmented generation. In *EACL: System Demonstrations*, 2024.
- [Google for Education, 2025] Google for Education. Transform teaching and learning with Gemini for education. <https://edu.google.com/ai/gemini-for-education/>, 2025. Accessed: 2026-01-29.
- [Güldal and Dinçer, 2025] Hakan Güldal and Emrah Oğuzhan Dinçer. Can rule-based educational chatbots be an acceptable alternative for students in higher education? *Education and Information Technologies*, 30(3):3979–4012, 2025.
- [Hunt, 2012] Tim Hunt. Moodle. *The Architecture of Open Source Applications*, 2, 2012.
- [Ji *et al.*, 2023] Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. Survey of hallucination in natural language generation. *ACM computing surveys*, 2023.
- [Kukla *et al.*, 2026] Michał Kukla, Sebastian Pergała, and Jan Skwarek. Didactic module for teaching and testing with conversational educational system. *BSc thesis, Warsaw University of Technology*, 2026.
- [Microsoft Learn, 2025] Microsoft Learn. Microsoft copilot in education. <https://learn.microsoft.com/en-us/microsoft-365/education/guide/1-reference/baseline-reference-copilot>, 2025. Accessed: 2026-01-29.
- [Moodle Project, 2025] Moodle Project. Plugin types - moodle developer resources, 2025. Accessed: 2026-01-24.
- [Motlagh *et al.*, 2023] Negin Yazdani Motlagh, Matin Khajavi, Abbas Sharifi, and Mohsen Ahmadi. The impact of artificial intelligence on the evolution of digital education: A comparative study of OpenAI text generation tools including ChatGPT, Bing Chat, Bard, and Ernie. *arXiv:2309.02029*, 2023.
- [Ostrowska *et al.*, 2026] Anna Ostrowska, Gabriela Majstrak, and Jan Opala. Learning module with conversational educational system. *BSc thesis, Warsaw University of Technology*, 2026.
- [Pérez *et al.*, 2020] José Quiroga Pérez, Thanasis Daradoumis, and Joan Manuel Marquès Puig. Re-discovering the use of chatbots in education: A systematic literature review. *Computer Applications in Engineering Education*, 28(6):1549–1565, 2020.
- [Quizlet, 2023] Quizlet. Introducing q-chat, the world's first ai tutor built with openai's chatgpt. <https://quizlet.com/blog/meet-q-chat>, 2023. Accessed: 2026-01-29.
- [StudyFetch, 2026] StudyFetch. Studyfetch: AI study guide and tool generator. <https://www.studyfetch.com/>, 2026. Accessed: 2026-01-23.
- [Valiathan, 2024] P. Valiathan. Bloom's taxonomy. In R. E. West and H. Leary, editors, *Foundations of Learning and Instructional Design Technology: Historical Roots & Current Trends*, pages 147–161. EdTech Books, 2024.
- [Wróblewska *et al.*, 2025] Anna Wróblewska, Bartosz Grabek, Jakub Świstak, and Daniel Dan. Evaluating LLM-generated Q&A test: A student-centered study. In *Artificial Intelligence in Education (AIED)*, Lecture Notes in Computer Science. Springer, Cham, 2025.

³However, the views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or the European Research Executive Agency. Neither the European Union nor the European Research Executive Agency can be held responsible for them.