SIMBA: creating chatbots for students tutoring

Gabriel Ferrettini[1], André C. A. Nascimento[2], Mar Pérez-Sanagustin[1], and Isabel Hilliger[3]

[1] Toulouse University, Institut de Recherche en Informatique de Toulouse
surname.name@irit.fr
https://www.irit.fr/
[2] Universidade Federal Rural de Pernambuco andre.camara@ufrpe.br
https://www.ufrpe.br/
[3] Pontificia Universidad Católica de Chile ihillige@uc.cl https://www.uc.cl/

Abstract. This demo paper presents SIMBA, a web-based tool designed to support instructors in creating learning activities with LLM-powered chatbots. It functions as a Socratic tutor, fostering students' metacognitive skills through structured questioning. By defining learning objectives and a sequence of guiding questions, instructors generate customized chatbots that act as learning companions. These can be shared with students, while instructors monitor interactions via an interactive analytics dashboard. SIMBA was evaluated during a workshop with eight instructors and the results suggest that SIMBA holds promise as a scaffolding tool for active learning, leveraging LLMs to enhance metacognitive engagement while providing instructors with actionable insights.

Keywords: Chatbot · Technology-enhanced learning · Artificial intelligence · Large Language Model · Higher Education

1 Teaching with LLM-based chatbot tutors

The integration of large language models (LLMs) in education presents both opportunities and challenges, as educators must balance AI's potential with pedagogical oversight [4]. While educational chatbots show promise in enhancing concept retention and self-efficacy [6], current systems primarily focus on student-facing interactions, leaving instructors without adequate tools to monitor or guide these learning experiences. This gap is particularly evident in the scarcity of teacher-facing dashboards, with only limited exceptions like StuGPTViz [2] beginning to explore this space.

To bridge this gap, we present SIMBA, a system that transforms student-LLM interactions into actionable pedagogical insights for instructors. Grounded in Prieto et al.'s orchestration framework [5], SIMBA provides educators with the monitoring and intervention capabilities needed to effectively integrate generative AI into teaching practices, reducing the current divide between AI-enhanced learning and evidence-based pedagogy.

2 SIMBA: Opening access to AI for teachers

SIMBA is a web-based instructional platform developed using Streamlit and integrated with the OpenAI API, designed to support educators in creating and managing pedagogically-grounded, LLM-powered chatbots. Grounded in Design-Based Research (DBR) methodology, the system was iteratively developed to address two core instructor needs [3]: (1) Activity Design: Intuitive interfaces for creating Socratic tutoring chatbots that guide students through structured questioning sequences; and (2) Interaction Analytics: A visual dashboard that transforms student-chatbot dialogues into actionable pedagogical insights. By bridging these capabilities, SIMBA supports the full lifecycle of AI-enhanced learning activities — from initial design by teachers to real-time monitoring of learning processes.

SIMBA's Activity Design interface offers teaching staff the possibility to create pedagogically structured, LLM-powered learning activities through an intuitive workflow. Instructors begin by defining core pedagogical elements: a descriptive objective that describes the purpose of the activity, a sequenced set of guiding questions to scaffold student inquiry, behavioural parameters including the chatbot's conversational tone (e.g., Socratic, supportive), and strictness rules governing answer disclosure. The system further supports curriculum alignment through document upload functionality, where course materials (readings, slides) are automatically processed to ground the chatbot's responses in domain-specific knowledge. All instructor inputs are automatically translated into optimized prompts for the OpenAI API, abstracting away technical complexities while preserving pedagogical intentionality in the resulting chatbot behaviour. Upon configuration, instructors receive a unique, shareable activity link (URL) that maintains configured constraints during student interactions, enabling seamless deployment of customized AI tutors. This shareable link serves as a secure access point for students, who encounter a purpose-built chatbot interface that rigorously adheres to the teacher's designed parameters — maintaining the specified Socratic tone, question sequencing, and answer disclosure policies. When students access the activity, they engage with an AI tutor that systematically fosters metacognitive reflection through the instructor-curated question flow, while remaining grounded in the uploaded course materials. This end-to-end design process transforms educators from passive consumers of generative AI into active designers of constrained learning experiences with LLMs.

All student-chatbot interactions are systematically logged and processed to generate SIMBA's Interaction Analytics Dashboard (figure 1, a visualization dashboard designed to transform raw dialogue data into pedagogical insights. The dashboard synthesizes students-LLM interactions patterns across three primary dimensions:

- General students' metrics: "conversation stats" tab. These metrics are mainly based on the students' number of messages exchanged with the chatbot and the length of those messages, along with the average conversation time. Those are aggregated data providing an overview of the students' engagement with each activity.

– Individual student metrics : "student stats" tab. Description ofetails of an individual student behaviour towards the activities, the number of activities consulted, their number of messages and their average length, along with the comparison of those metrics with those of the other students.
– raw conversational data : "raw data" tab. Allows the teacher to see the conversation of a student with the chatbot.

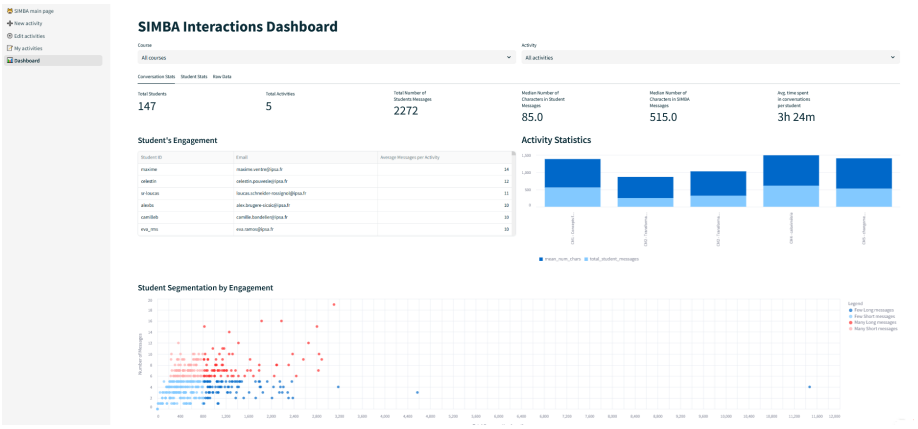

Fig. 1. View of SIMBA's dashboard "conversation stats"

3 Workshop: Presenting SIMBA to teachers

3.1 Workshop description

To evaluate SIMBA's core functionalities, we conducted a workshop with x instructors from Computer Science and Psychology disciplines at a European university. The study addressed three research questions:

– **RQ1**: To what extent do instructors perceive the Activity Design interface as usable for creating LLM-powered learning activities?
– **RQ2**: To what extent can instructors effectively navigate, interpret, and derive actionable pedagogical insights from SIMBA's Interaction Analytics Dashboard?
– **RQ3**: In what ways would educators envision integrating SIMBA into their existing teaching workflows?

SIMBA was preloaded with interaction data from a first-year Thermodynamics course experiment, allowing participants to explore pre-built activities. The 60-minute workshop comprised: **Phase 1 (Presentation)**: 5-minute introduction to SIMBA's capabilities. **Phase 2 (Activity Design)**: Participants created an activity based on a provided scenario. And **Phase 3 (Dashboard**

Exploration): Guided analysis of the Interaction Analytics Dashboard using targeted questions (e.g., "Which student showed the highest engagement with the LLM?"). Workshop materials are available at this link. We employed two validated instruments for data gathering: (1) A post-Phase 2 **usability survey** using a 5-point Likert scale adaptation of the System Usability Scale ([1]), and (2) A post-Phase 3 **understandability assessment** featuring: (1) Guided discovery tasks with concrete dashboard queries, and (2) A validated interpretability survey [blinded for review] The collected data were independently analysed by two researchers using qualitative methods. Following an inductive coding approach, we systematically examined the dataset to identify emerging themes and patterns relevant to each research question using Dedoose. Emergent codes were used to tag the participant responses. Those are summarized here with the number of participants for which the code occurred at least once.

Concerning the dashboard, 1 participant proposed to use the dashboard' information to monitor the chatbot answers, 5 to monitor the student's understanding of the course, 1 to prepare for future courses, and 1 to verify the students' engagement with SIMBA activities. 1 participant expressed a negative opinion toward the usability of the dashboard while 4 expressed a positive opinion, 1 participant specifically cited a dashboard negative feature while 6 cited positives ones. Finally, 3 participants expressed a negative opinion toward their understanding of the dashboard, while 3 expressed a positive one. 1 participant proposed to Use SIMBA's conversations to correct the students' misconceptions, 5 to let students work autonomously, 2 to identify students weaknesses to be corrected and 5 to help students revise their courses. Concerning the activities editing, 3 participants expressed a negative opinion toward its usability, 4 expressed a positive opinion, 1 cited specifically a negative feature, while 5 cited positive ones.

The analysis was done with all primary data analysis in French, the native language of participants. For results presentation, we carefully selected and translated representative quotes from the original French dataset.

3.2 Results and discussion

RQ1. Activity Design Usability The SUS responses for activity creation indicated strong usability, with most participants agreeing to positive statements and disagreeing with negative ones. A notable exception was the statement "Most people will learn to use SIMBA very quickly", which received mixed responses (four neutral, three slightly agree, and one disagree). All participants successfully created and tested activities independently. When identifying positive aspects, teachers highlighted both specific editing tools ("The interface is rather user-friendly") and the general interface design (*"I like being able to ask specific questions from the course without necessarily needing to prepare an answer"*). Negative feedback focused primarily on editing features, including one participant's difficulty understanding the setup process (*"Overall, I'm finding it hard to understand the tool..."*) or desire for greater control (*"More control over the answers to the questions"*).

Overall, SIMBA's editing section proved generally user-friendly but presented challenges for some users. While most criticism addressed technical limitations like interface responsiveness or answer control, the activity creation functionality was broadly well-received as pedagogically valuable. The positive reception suggests SIMBA succeeds as a practical tool, despite areas needing refinement.

RQ2. Usability and understandability for the Dashboard Teachers identified several strengths of the dashboard, with three specifically praising its features. One highlighted the usefulness of *"The graphs comparing the student's use of SIMBA with that of other students"*. All participants accurately interpreted and located information during comprehension tasks. However, one teacher emphasized the need for more student feedback mechanisms, particularly regarding AI hallucinations and usage patterns. Survey responses revealed mixed perceptions of understandability: teachers criticized the lack of clarity in indicator explanations (e.g., *"The dashboard provides sufficient explanations of how the various elements relate to each other"*) but strongly affirmed its instructional value (e.g., *"The dashboard provides information that helps me to guide my students in organizing their learning process"*). Overall, the dashboard was seen as comprehensible and educationally useful, though two key improvements emerged: clearer visualizations and explanations, and more detailed student interaction data to enhance classroom utility. Despite these areas for refinement, results support the dashboard's effectiveness in aiding teaching goals. Participants also proposed several pedagogical uses for the analytics. Five instructors described using it to diagnose student understanding—one explained: *"[I would use the dashboard] to see what students don't understand in my lessons and therefore clarify these points, for example."* Other suggested applications included evaluating chatbot effectiveness, refining course content, and tracking engagement patterns. All participants articulated concrete strategies for applying the data, highlighting its perceived value for formative assessment and instructional design. While findings show promise for classroom integration, longitudinal research is needed to confirm long-term efficacy. The dashboard clearly translates interaction data into meaningful insights, though its sustained impact on practice requires further study.

RQ3. Integration of SIMBA in teaching and learning workflows Participants proposed a range of pedagogical uses for SIMBA, highlighting its versatility as an instructional tool. Four teachers suggested using it to assign autonomous exercises, and another four saw it as a valuable tutor for student revision—prompting activities like summarizing course content and developing answers through questioning. One participant outlined a dual-activity strategy: *"I'm thinking of creating two different activities for each course, [one] where the student can ask questions [...], and another activity where the chatbot asks course questions [and exercises] linked to the lesson."* Another emphasized its diagnostic use: *"I could use it to help the student identify where their gaps are regarding the course."*

Importantly, all participants spontaneously proposed concrete implementation ideas, signalling strong perceived utility. These findings suggest SIMBA addresses key instructional needs: (1) differentiated instruction through autonomous tasks, (2) metacognitive development via structured revision, and (3) formative assessment through gap identification. While initial feedback is promising, future research should explore the impact of these uses on learning outcomes and instructional practice.

4 Conclusion

This demo paper has presented SIMBA, a tool to support teachers in the design, deployment and monitoring of LLM-based activities. Our findings suggest that by eliminating the need for LLM training or manual prompt engineering, SIMBA successfully enables educators to create and deploy pedagogical chatbots for student tutoring—particularly for course revision and supplementary assistance. The integrated analytics dashboard, which provides actionable insights into student-chatbot interactions, further empowers teachers to identify challenging course concepts and monitor learning progress. Participant enthusiasm during our workshop indicates that SIMBA represents a promising approach for implementing and studying generative AI in authentic educational settings. However, future research should expand these preliminary findings through longitudinal classroom deployments and more diverse teacher cohorts to fully establish the system's educational benefits and practical implementation requirements.

References

1. Brooke, J., et al.: Sus-a quick and dirty usability scale. Usability evaluation in industry **189**(194), 4–7 (1996)
2. Chen, Z., Wang, J., Xia, M., Shigyo, K., Liu, D., Zhang, R., Qu, H.: Stugptviz: A visual analytics approach to understand student-chatgpt interactions. IEEE Transactions on Visualization and Computer Graphics (2024)
3. Hillinger, I., Carlos González, M.P.S., Ferrettini, G., Esteban Villalobos, Andre Nascimento, S.C.: From one-way text from socratic dialogue: Designing an artificial intelligence chatbot to support student learning in higher education. 15th International Conference on Learning Analytics and Knowledge (LAK25) (2025)
4. Kadaruddin, K.: Empowering education through generative ai: Innovative instructional strategies for tomorrow's learners. International Journal of Business, Law, and Education **4**(2), 618–625 (2023)
5. Prieto, L.P., Martínez-Maldonado, R., Dimitriadis, Y.: Orchestrating learning analytics (orla): Supporting inter-stakeholder communication about adoption of learning analytics at the classroom level. IEEE Transactions on Learning Technologies **12**(3), 331–345 (2018)
6. Rudnik, Y.: The use of artificial intelligence chatbots in teaching foreign languages as an innovative interactive technology. Educological discourse **45**(2) (2024)