Mobile-Accessible Seamless Agent (MASA): Using ChatbotLLM for Grounding Chatbots in Teacher-Provided Materials

Fuzheng ZHAO[a*], Zibusiso MAFAITI[b], Yongjin LI[b] & Maiga CHANG[b]
[a]*Education Technology Center, Jilin University, China*
[b]*School of Computing and Information Systems, Athabasca University, Canada*
* zhaofz635@gmail.com

Abstract: One of the greatest challenges facing AI-driven educational chatbots is the issue of hallucinations. At its core, factual errors, logical inconsistencies, and fabricated content severely undermine the foundational principles of truth and rigor in education. While current research methods have reduced the incidence of hallucinations, they remain susceptible to retrieval noise and lack fine-grained teacher control due to their black-box nature. To address these limitations, this study adopts ChatbotLLM to develop a browser-based educational chatbot that is a Mobile-Accessible Seamless Agent (MASA) to minimize adoption barriers. At the end of this paper, the study proposes a comparative experimental plan to validate significantly reduced hallucination rates in knowledge comprehension, citation-oriented, and reasoning tasks. By prioritizing teacher-led control and curriculum fidelity, ChatbotLLM and MASA advance human-machine interaction models for trustworthy AI in education.

Keywords: Hallucination mitigation; educational chatbot; teacher-grounded retrieval; configurable ai control; mobile accessible

1. Introduction

Led by large language models such as ChatGPT, conversational AI bots have rapidly gained traction in higher education (Yan et al., 2024). With core features such as natural language multi-turn conversations, personalized adaptation, and real-time feedback (Lin, Huang & Yang, 2023), these tools simulate continuous interaction with human mentors, supporting student inquiries and becoming indispensable teaching aids (Zhao et al., 2023). However, these advantages come with significant drawbacks, most notably the hallucination issues (Ji et al., 2023). Simply put, AI may generate content that appears plausible yet contains factual errors, computational biases, fabricated citations, and logical contradictions (Karpathy, 2015).

Hallucinations cannot be eliminated, which are an inevitable outcome of model architecture design and trade-offs. Consequently, current research focuses on mitigating significant risks (Singh et al., 2025), with prompt engineering, fine-tuning, and grounding emerging as the three primary strategies (Ji et al., 2023; Joshi, 2025). As generative AI applications deepen in education, the hybrid layered approach has become one of the most prominent methods. This approach's defining characteristics include shifting from pure internal parameter optimization to external knowledge grounding, process supervision, and self-consistency (Lewis et al., 2020). It significantly reduces hallucination rates, particularly knowledge-based hallucinations and factual errors (Lewis et al., 2020; Wei et al., 2022), while enhancing interpretability and controllability. (Wang et al., 2022).

To address hallucination issues in conversational AI agents, De Silva and Chang (2025) design ChatbotLLM that adopts sentence transformer and cosine similarity to find question related paragraphs for users include client apps to choose the top N question related paragraphs, to tweak threshold of cosine similarity value, to opt in/out the use of large language model (i.e., Google Gemini) for response rephrasing and/or RAG. This study

presents a Mobile-Accessible Seamless Agent (a.k.a. MASA) that is a ChatbotLLM-enabled browser-based app for providing learners a straightforward and intuitive user experience.

2. Hallucination

2.1 Definition, Cause, and Impact

The fundamental causes of hallucination issues primarily stem from the statistical nature of parametric knowledge and the generalization-hallucination tradeoff (Zhang et al. 2023). Nah et al. (2023) analyzed the underlying mechanisms of hallucination generation. The statistical nature of parametric knowledge means that large language models are not databases. Instead, they learn the next most probable word by processing vast amounts of text, prioritizing fluency and coherence. Conversely, the generalization-hallucination tradeoff primarily stems from the requirement for robust generalization capabilities. Overly suppressing inference reduces this capability, making hallucinations an inevitable trade-off for pursuing generality and fluency.

This issue gained significant exposure following ChatGPT's emergence, prompting systematic academic research. Education, however, relies on accurate knowledge foundations, rigorous logical reasoning, and reliable evidence citation (Willingham, 2008; Abrami et al., 2015). Systematic research by Kasneci et al. (2023) indicates that when these foundations are undermined by hallucinations, students may develop erroneous conceptual frameworks, experience reduced learning motivation, and suffer other irreversible learning impairments.

2.2 Limitations and Trends

It is widely acknowledged that eliminating hallucination issues remains unfeasible, with current research prioritizing the mitigation of significant risks (Singh et al., 2025). According to Joshi's (2025) analysis, prompt engineering, fine-tuning, and grounding have emerged as the three primary strategies for alleviating hallucination issues. Concurrently, as generative AI educational applications like ChatGPT gain traction, the hybrid layered approach has emerged as one of the most prominent methods. Lewis et al. (2020) observed a shift in the evolution of hallucination mitigation techniques. The hybrid layered approach, characterized by external knowledge grounding, process supervision, and self-consistency, is gradually replacing purely internal parameter optimization methods. With this approach, hallucination rates have been significantly reduced, particularly for knowledge-based hallucinations and factual errors (Lewis et al., 2020; Wei et al., 2022). Additionally, models have gained enhanced interpretability and controllability (Wang et al., 2022).

The key to developing a hybrid layered approach. The latest trend in hallucination mitigation involves shifting from internal parameter optimization toward hybrid strategies combining external knowledge anchoring with human-controllable configuration (Kasneci et al., 2023; Yan et al., 2025). Particularly in educational settings, teacher-curated authentic materials as the sole knowledge source, coupled with configurable switches to control generative behavior, are considered the key pathway to address the lack of teacher control and residual generative bias.

2.3 ChatbotLLM

ChatbotLLM[1] (De Silva et al., 2025) is IEEE Northern Canada Section's Capstone Project Award winner project, is a system to automatically train chatbots on the materials uploaded by users. ChatbotLLM is not only capable of learning from materials in traditional formats like PDF, PPT(X), DOC(X), and TXT, but also can read XLS(X), image-scanned based PDF, images in PDF, DOC(X), and PPTX, and even MP3 with the help of Google's Speech

[1] https://chatbot.vipresearch.ca

Recognition library. When a course is created and the course materials are uploaded, ChatbotLLM will have five modules, convert-to-text, extract-urls, crawl-and-extract, enrich, and data preparation, working together to prepare the dataset for chatbot training.

ChatbotLLM follows a three-tier architecture consisting of Presentation (Client App), Orchestration (API Endpoints), and Logic (Service Provider) as shown in Figure 1. The client app (MASA) captures the user's question, the selected course, and model configuration. The API endpoints validate and route these requests. The service provider encodes the query, performs vector-based retrieval using cosine similarity across a precomputed index of lecturer-uploaded materials, and returns the top-ranked segments/paragraphs. An optional constrained generative model may rephrase retrieved content for readability while preserving factual grounding.

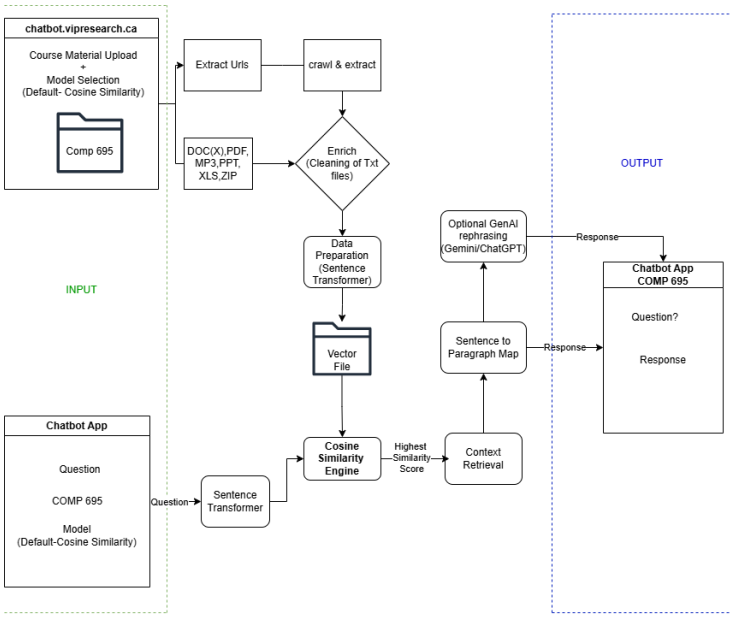


Figure 1. *ChatbotLLM and its client app architecture*

A learner can interact with ChatbotLLM and ask course related questions via a client app, for example, a LINE bot or a Discord bot. The client app will forward the learner's question to ChatbotLLM service provider. When ChatbotLLM service provider receives the request package, it first adopts sentence transformer and cosine similarity to search and identify the top N question related paragraphs from the teacher uploaded materials based on the pre-set cosine similarity threshold and top N value. According to teacher's setting (opt-out or not), ChatbotLLM will then use (or not use) large language model (Google Gemini) for either rephrasing and/or RAG before sending the response back to client app.

Different from other Generative AI tools or systems, all the responses were identified and generated by ChatbotLLM are based on the materials teachers provided instead of generating things based on everything it learned from Internet or other sources and trying to answer everything including hallucinations for its users include learners and client apps. So ChatbotLLM would tell users "NO, I don't know" if it cannot find any relevant content for user's question from the given materials uploaded.

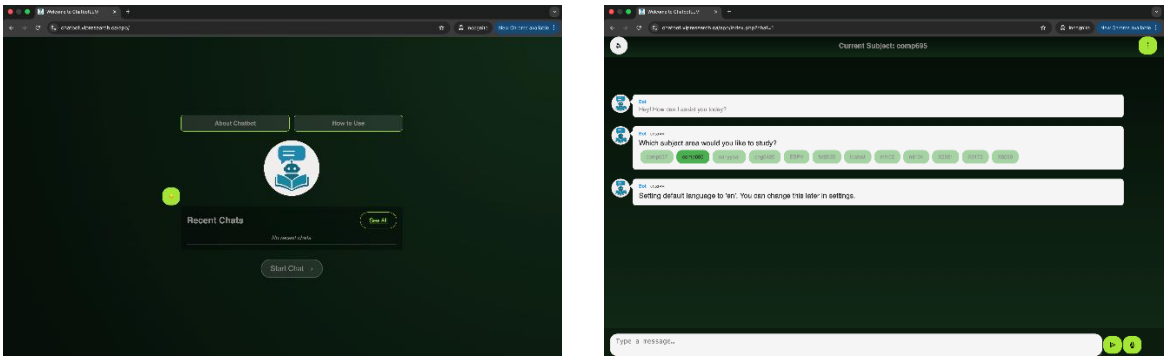
3. Mobile-Accessible Seamless Agent (MASA)

To achieve Seamless Accessibility objective, the research team prioritizes frictionless user experience with mobile-accessible and simplicity design and develops Mobile-Accessible Seamless Agent (MASA[2]). Removing authentication barriers such as complex logins ensures instant access to learning materials on students' devices via a familiar, intuitive interface.

[2] https://chatbot.vipresearch.ca/app

MASA's UI/UX follows common mobile app patterns to minimize cognitive load and support natural, companion-like interaction between learners and the Chatbot service provider. Research shows these interactions improve learning outcomes through the Persona, Proteus, and Protégé effects (Pérez-Marín, 2021). By positioning the chatbot as an approachable peer rather than a rigid tool, MASA strengthens student engagement and retention. This seamless accessibility effectively delivers the ChatbotLLM's core hallucination-mitigation mechanisms including teacher-grounded retrieval pipeline and configurable generative control.

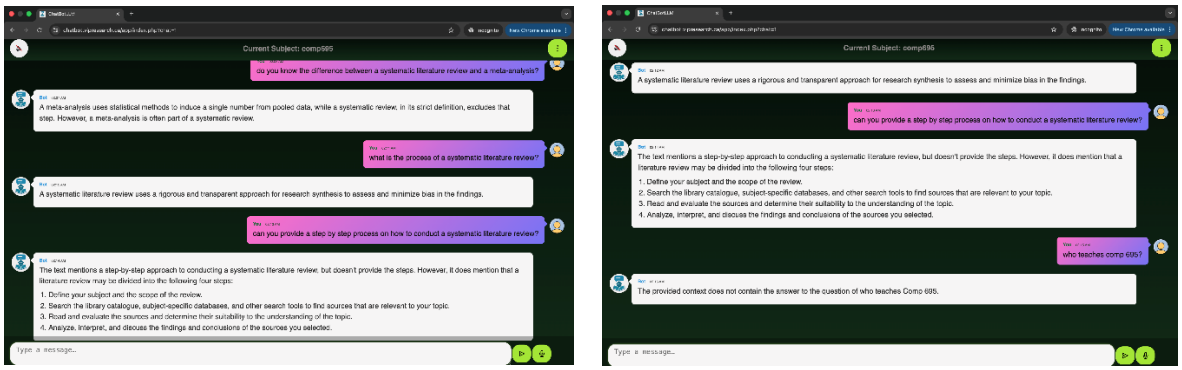
When learners are told to use MASA for finding answers of course material related questions, they can see a clean mobile accessible landing page as Figure 2(a) shown below. Besides starting the chat immediately, learners can also see the latest conversations with MASA on the landing page. As soon as they click "Start Chart →" button, they will be greeted and asked which subject they would like MASA to assist. Figure 2(b) shows the subjects available to the public – even when a subject is open to the public, teachers can still configure it by setup an access key to protect their uploaded content from learners other than their students.



(a) MASA's landing page *(b) subjects offered and selected*

Figure 2. MASA, the browser-based client app

As soon as a subject is chosen, learners can start asking any teaching materials related questions. Figure 3(a) shows that they not only can use keyboard to enter their questions but also can use microphone to "talk" to MASA with its speech-to-text function's help – just like what they can do within commercial social messaging apps (e.g., Facebook Messager, Google Chat, LINE, WeChat, WhatsApp, etc.). Similarly, MASA can also talk back due to it having built-in text-to-speech function. Of course, if learners don't want to "hear" from MASA but only see the text MASA typing, then they can click the speaker icon at top-right hand side on the screen to mute (or unmute) MASA. To avoid hallucination, MASA will not attempt to answer any question it cannot find correspondent content from the uploaded materials teachers uploaded as Figure 3(b) shown.



(a) asking questions with keyboard/mic *(b) MASA will not attempt to answer*

Figure 3. Asking MASA teaching material (non-)related question

4. Evaluation Plan

To validate the effectiveness of the educational chatbot proposed for mitigating hallucination issues, a comparative experiment was designed. The experiment aims to verify a strong knowledge anchoring mechanism based on teacher-approved textbooks and academic literature. The experiment examines whether the mechanism can significantly reduce the risk of hallucinations in educational applications of conversational AI without compromising learning outcomes. This experiment plans to recruit 60 university students as participants, all from the same or similar disciplinary backgrounds to minimize domain knowledge differences affecting results. Participants will be randomly assigned to either an experimental group or a control group. The experimental group uses MASA with ChatbotLLM service provider proposed in this study. This chatbot permits responses only based on textbooks and academic literature uploaded and verified by instructors. The control group uses MASA with a large language model (LLM) with unrestricted knowledge sources.

Learning activities will be selected from the university *Introduction to Artificial Intelligence Technology*, intentionally designing three task types prone to inducing hallucinations: knowledge comprehension tasks, evidence and citation-oriented tasks, and reasoning-based tasks (Ji et al., 2023). Knowledge comprehension tasks will require participants to ask the questions about the concept of machine learning from the textbook to verify the factual accuracy and conceptual consistency of the chatbot's responses. The tasks are completed through free-form dialogue to mirror how students use a large language model AI in real learning contexts.

The experimental process will comprise four stages: pre-test, chatbot usage, post-test, and questionnaire survey. Before the experiment begins, all participants will complete a 10-minute pre-test to assess their prior knowledge of relevant learning content, ensuring comparability between the experimental and control groups at baseline. Subsequently, participants from both groups entered the chatbot usage phase, where they will complete assigned learning tasks using their respective chatbot via two different links. This learning activity will last 30 minutes. Throughout this process, the chatbot automatically record complete interaction logs, including student questions, chatbot responses, and corresponding timestamps. After the experiment concluded, all participants will complete a 10-minute post-test of equivalent difficulty to the pre-test but with non-repetitive questions to evaluate changes in learning outcomes. Finally, participants will complete questionnaires. The study will employ Gray and DiLoreto's (2016) satisfaction scale to subjectively evaluate perceptions of the chatbot's technical output and overall satisfaction.

The core focus of the experimental evaluation metrics is the effectiveness of hallucination mitigation, while also considering learning outcomes. First, an objective hallucination assessment metric will be constructed based on the chatbot's dialogue logs. This study plans to invite two domain experts with relevant disciplinary backgrounds to independently annotate the chatbot-generated responses, judging whether they contain factual errors, fabricated or misattributed information, or logical inconsistencies, all indicative of hallucination phenomena. Any response exhibiting any of these issues will be deemed to contain a hallucination. Building on this, the study will calculate differences between the two chatbots in overall hallucination incidence and distribution across hallucination types. A graded severity scale will further assess potential risks to learning. Additionally, to prevent hallucination mitigation from compromising learning outcomes, the experiment will compare pre- and post-test performance differences between participant groups to evaluate the chatbot's impact on learning effectiveness.

5. Conclusion

In the field of education, factual accuracy, logical rigor, and reliable citations form the foundation of effective learning (Kasneci et al., 2023; Ji et al., 2023). Hallucination remains an inherent challenge in educational applications of large language model-based chatbots. ChatbotLLM strictly adheres to a single source of truth derived exclusively from teacher-

uploaded and verified authentic textbooks or academic literature, eliminating reliance on parameterized knowledge and reducing retrieval noise at its source. This paper presents the MASA (Mobile-Accessible Seamless Agent), a client app of ChatbotLLM that has hallucination mitigation implemented. MASA is a browser-based instructional dialogue chatbot that advances the emerging trend of teacher-led instructional foundations and configurable safety layers (Kasneci et al., 2023; Yan et al., 2025).

A limitation of this study is the lack of empirical validation. To address this, we plan to conduct a comparative experiment to empirically validate the effectiveness of these mechanisms. The study will recruit 60 university students to use MASA and compare ChatbotLLM against an unrestricted LLM baseline system on hallucination-prone task types. This will quantify reductions in hallucination occurrence rates, severity, and distribution patterns while ensuring learning outcomes remain unaffected.

References

Abrami, P. C., Bernard, R. M., Borokhovski, E., Waddington, D. I., Wade, C. A., & Persson, T. (2015). Strategies for teaching students to think critically: A meta-analysis. *Review of educational research*, *85*(2), 275-314.

De Silva, S., & Chang, M. (2025). ChatbotLLM – Training Educational Chatbots on the Materials Uploaded by Teachers. In: the Proceedings of 13th International Conference on Information and Education Technology (ICIET 2025), Fukuyama, Japan, April 18-20, 2025. https://doi.org/10.1109/ICIET66371.2025.11046253

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., ... & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM computing surveys*, *55*(12), 1-38.

Joshi, S. (2025). Mitigating LLM Hallucinations: A Comprehensive Review of Techniques and Architectures. *Available at SSRN 5267540*.

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and individual differences*, *103*, 102274.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in neural information processing systems*, *33*, 9459-9474.

Lin, C.C., Huang, A.Y.Q., Yang, S.J.H. (2023), A Review of AI-Driven Conversational Chatbots Implementation Methodologies and Challenges (1999–2022), *Sustainability, 15*(5), 4012

Nah, F., Zheng, R., Cai, J., Siau, K., & Chen, L. (2023). Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. *Journal of information technology case and application research*, *25*(3), 277-304.

Singh, S., Saha, R., Kumar, G., Nayyar, A., & Kim, T. K. (2025). Are you hallucinated? Insights into large language models. *ICT Express*.

Wang, X., Wei, J., Schuurmans, D., Le, Q., Chi, E., Narang, S., ... & Zhou, D. (2022). Self-consistency improves chain of thought reasoning in language models. *arXiv preprint arXiv:2203.11171*.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., ... & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, *35*, 24824-24837.

Willingham, D. T. (2008). Critical thinking: Why is it so hard to teach?. *Arts Education Policy Review*, *109*(4), 21-32.

Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., ... & Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, *55*(1), 90-112.

Zhang, Y., Li, Y., Cui, L., Cai, D., Liu, L., Fu, T., ... & Shi, S. (2025). Siren's Song in the AI Ocean: A Survey on Hallucination in Large Language Models. *Computational Linguistics*, *51*(4), 1373-1418.

Zhao, F., Liu, G. Z., Zhou, J., & Yin, C. (2023). A learning analytics framework based on human-centered artificial intelligence for identifying the optimal learning strategy to intervene in learning behavior. *Educational Technology & Society*, *26*(1), 132-146.