Evaluating a GPT-4 and Retrieval-Augmented Generation-Based Conversational Agent to Enhance Learning Experience in a MOOC

Fatma Miladi[1], Valéry Psyché[1], Awa Diattara[2], Nour El Mawas[3] and Daniel Lemire[1]
[1]*TELUQ University, 5800 rue Saint-Denis, Montreal, QC H2S 3L5, Canada*
[2]*LANI, Gaston Berger University, Saint-Louis, Senegal*
[3]*Université de Lorraine, Crem, F-57000 Metz, France*
{*fatma.miladi, valery.psyche, daniel.lemire*}*@teluq.ca, awa.diattara@ugb.edu.sn, nour.el-mawas@univ-lorraine.fr*

Keywords: Conversational Agent, Educational Chatbot, Generative AI, Retrieval-Augmented Generation (RAG), Massive Open Online Course (MOOC), Artificial Intelligence in Education.

Abstract: Massive Open Online Courses (MOOCs) face significant challenges due to low completion rates, primarily caused by insufficient personalized support for learners. To address this, we developed a pedagogical AI-powered conversational agent enhanced with Retrieval-Augmented Generation (RAG) to provide real-time, contextually relevant support. Our evaluation with 25 learners demonstrated a statistically significant knowledge gain in the experimental group compared to the control group. Additionally, the agent achieved a high System Usability Scale (SUS) score. These findings highlight the potential of AI technologies to enhance online learning environments and inform future research on their role as learning companions in distance education.

1 INTRODUCTION

Massive Open Online Courses (MOOCs) allow students worldwide to learn at their own pace and on flexible schedules. This flexibility has contributed to the rapid growth in the popularity of MOOCs. However, despite high enrollment rates, the completion rate of MOOCs remains low. On average, less than 10% of learners complete a MOOC (Yin et al., 2019), raising concerns about the effectiveness of these courses in terms of learner retention and success. One of the key challenges contributing to these low completion rates is the lack of personalized support during the online learning course, which is crucial for learner retention and success.

A significant issue is the lack of instructor feedback in online courses, which leaves learners without the guidance they need to stay motivated and engaged in their learning. This absence of direct interaction, combined with limited opportunities for teamwork or group interaction, contributes to learner demotivation and lower retention rates (Hone and El Said, 2016).

Although MOOCs typically include features such as discussion forums to facilitate social interaction among learners, participation remains low, with only 5% to 12% of learners actively engaging in these discussions (Chiu and Hew, 2018). Additionally, the instructor's involvement in these forums is often minimal, leaving many learners without timely support. This challenge is further complicated by the fact that many participants feel unsure how to initiate meaningful conversations and may be hesitant or shy to engage.

Generative Artificial Intelligence (GAI) has emerged as a promising solution to these challenges. Specifically, models based on Generative Pre-trained Transformers (GPTs) leverage vast amounts of data to generate human-like text responses. These technologies are increasingly being used in various settings, including education (Adeshola and Adepoju, 2024; Mariani et al., 2023). However, despite its potential, research on the application of Generative AI in education, particularly in the context of MOOCs, is still in its early stages (Chiu, 2024).

To bridge this gap, we designed and implemented a pedagogical conversational agent leveraging GPT with Retrieval-Augmented Generation (RAG). This integration enables the agent to deliver contextually accurate and course-specific responses by retrieving information from a database of documents used in the course design. This capability aims to enhance knowledge acquisition and foster a supportive learning environment by providing relevant and precise information in real time. Specifically, we address the

Miladi, F., Psyché, V., Diattara, A., El Mawas, N. and Lemire, D.
Evaluating a GPT-4 and Retrieval-Augmented Generation-Based Conversational Agent to Enhance Learning Experience in a MOOC.
DOI: 10.5220/0013366100003932

In *Proceedings of the 17th International Conference on Computer Supported Education (CSEDU 2025) - Volume 2*, pages 347-354
ISBN: 978-989-758-746-7; ISSN: 2184-5026

following research questions:

RQ1: Does the use of a GPT and RAG-enhanced conversational agent alongside learners in the MOOC affect their knowledge acquisition?

RQ2: Can a conversational agent enhanced by GPT and RAG fulfill learners' expectations in terms of usability?

This paper is structured as follows: Section 2 provides a literature review on chatbots powered by LLMs and RAG in education. Section 3 presents the design of the conversational agent. Section 4 presents the research methodology. Section 5 details the results of the quantitative analyses. Finally, Section 6 discusses the findings, and Section 7 concludes the paper with implications for future research.

2 LITERATURE REVIEW

This section provides an overview of LLM-based conversational agents in education, highlighting their benefits and challenges. It then introduces the RAG approach and examines how it improves the factual accuracy and contextual relevance of chatbot responses in educational settings.

2.1 LLM-Based Conversational Agents in Education

The emergence of LLMs, such as ChatGPT, has significantly enhanced educational tools by providing richer, more adaptive interactions tailored to diverse learner needs. Abdelghani et al. (2022) demonstrated that GPT-3 fosters critical thinking in children by generating learning hints, which stimulate curiosity and improve knowledge retention. Similarly, Xie et al. (2024) found that LLM-based chatbots enhance autonomy for learners seeking social interaction. However, for those focused on knowledge acquisition, frequent interactions may reduce autonomy. This highlights the need to balance emotional support and cognitive guidance for effective learning.

Despite these advantages, LLMs face challenges, particularly their tendency to generate incorrect or biased information, known as hallucinations (Ji et al., 2023). In educational settings, such errors can mislead learners and compromise learning quality. To mitigate this issue, Retrieval-Augmented Generation can improve accuracy by retrieving relevant external information, reducing hallucinations, and enhancing response reliability (Shuster et al., 2021).

2.2 Theoretical Foundations of RAG

RAG, introduced by Lewis et al. (2020), enhances LLM reliability by integrating external knowledge retrieval into the generation process. It follows three main stages: indexing, retrieval, and generation (Gao et al., 2023).

In the indexing stage, text from various sources is processed and transformed into numerical vector representations using an embedding model. These vectors encode the semantic meaning of the text, enabling the system to efficiently organize and store information in a database for retrieval.

The retrieval stage begins when a user submits a query. The system converts the query into a vector representation using the same embedding model applied during indexing. It then compares this vector with stored vectors, identifying the most relevant text sections based on similarity scores.

In the generation stage, the retrieved text sections are combined with the user's query to form a context-enriched prompt. This prompt is then processed by an LLM, which generates a response that is more accurate and contextually relevant.

2.3 RAG-Based Conversational Agents in Education

Recent advancements in RAG have shown a promising ability to improve the accuracy and relevance of chatbot responses in education. Taneja et al. (2024) introduced Jill Watson, a virtual teaching assistant that uses RAG to retrieve relevant course materials, thereby reducing hallucinations and enhancing response quality. The study compared Jill Watson to virtual assistants not enhanced by RAG, demonstrating a clear improvement in response quality and a reduction in errors. Similarly, Yan et al. (2024) demonstrated how the chatbot VizChat uses RAG to enhance learning analytics dashboards, providing accurate and transparent explanations of visual data, reducing errors, and improving user comprehension. Likewise, Liu et al. (2024) developed CS50 Duck, a GPT-4-based conversational agent enhanced with RAG to support students in the course. It outperformed ChatGPT alone by providing more accurate and course-relevant responses. In parallel, Wang et al. (2023) developed ChatEd, a conversational agent for higher education that combines contextual information retrieval with ChatGPT. Its evaluation focused on relevance, accuracy, and usefulness. Compared to ChatGPT alone, ChatEd performed better on these criteria by leveraging a contextual database to align responses with course content. Likewise, Miladi et al. (2024)

examined the impact of RAG integration in GPT-4 and GPT-3.5 on response accuracy in an AI MOOC. Their findings showed that RAG-enhanced models outperformed their standard counterparts.

However, despite these promising advancements, current research primarily focuses on technical metrics such as accuracy, contextual relevance, and response clarity. These studies often overlook an in-depth exploration of the direct impact of RAG-enhanced language models on learning in real educational environments, such as MOOCs. Our study addresses this gap by evaluating the effect of a RAG-enhanced agent on learners' knowledge acquisition and usability.

3 MODEL DESIGN

We designed a conversational agent model based on the RAG technique (Gao et al., 2023) integrated with GPT-4. The model aims to enhance user interaction by combining the retrieval of relevant information from a specialized database with the generative capabilities of large language models. Figure 1 illustrates the architecture of our GPT-RAG conversational agent, which consists of seven key stages.

1. **Collection and Standardization of Documents** (Figure 1 (a))**.** We extracted documents from the MOOC on artificial intelligence (Psyché, 2020) as the primary source of information, including explanatory texts, video transcripts, and tables. These sources were converted into a uniform plain text format to ensure consistency for further processing.
2. **Document Segmentation** (Figure 1 (b))**.** The preprocessed documents were divided into smaller segments using Langchain's recursive character-based text splitter. Each segment was set to 2000 characters with a 200-character overlap to maintain context, following the parameters defined by Aymeric Roucher[1].
3. **Embedding Model** (Figure 1 (c))**.** The segmented text was transformed into numerical representations, called embeddings, using OpenAI's `text-embedding-ada-002` model (Neelakantan et al., 2022). These embeddings capture the meaning of the text, allowing the system to find relevant information based on similarity in meaning rather than just matching words.
4. **Knowledge Base** (Figure 1 (d))**.** The generated embeddings were stored in a structured knowledge base. This enables the system to retrieve relevant information efficiently when a learner asks a question.
5. **Query Processing** (Figure 1 (e))**.** When a learner submits a question, it is transformed into an embedding vector using the same embedding model as in stage (c). This transformation allows the system to compare the meaning of the question with the stored information in the Knowledge Base, even if the exact words do not match.
6. **Semantic Search** (Figure 1 (f))**.** The system compares the numerical representation of the question with the stored vectors using cosine similarity (Vijaymeena and Kavitha, 2016). It then selects the three most relevant text segments to provide context for generating a response.
7. **Enriched Prompt and Response Generation with GPT-4** (Figure 1 (g))**.** The selected text segments are combined with the original question to create an enriched prompt, which is then sent to GPT-4. This ensures that the response is based on reliable sources, which can help reduce errors and enhance accuracy and contextual relevance.

4 RESEARCH METHODOLOGY

Our research is based on a MOOC focused on artificial intelligence (Psyché, 2020). The course is structured into four modules, each covering different aspects of AI: general AI concepts, symbolic AI, connectionist AI, and AI applications in education. This study concentrates specifically on the first module.

We employed a quantitative data collection technique to address the research questions. Data were gathered through questionnaires and analysed using descriptive statistics to answer RQ1 and RQ2. This approach was selected to provide a clear overview of the data and support the analysis of experimental outcomes, thereby improving the study's replicability.

Ethical considerations were a key aspect of this study. To ensure data privacy, access to collected data was restricted to authorized personnel only. All participants provided informed consent, and the study received approval from TELUQ University's Ethics Committee (approval no. 10/2023).

4.1 Research Participants

The present study involved a sample of master's and bachelor's degree students in Informatics at a public

[1] https://huggingface.co/learn/cookbook/en/rag_evaluation

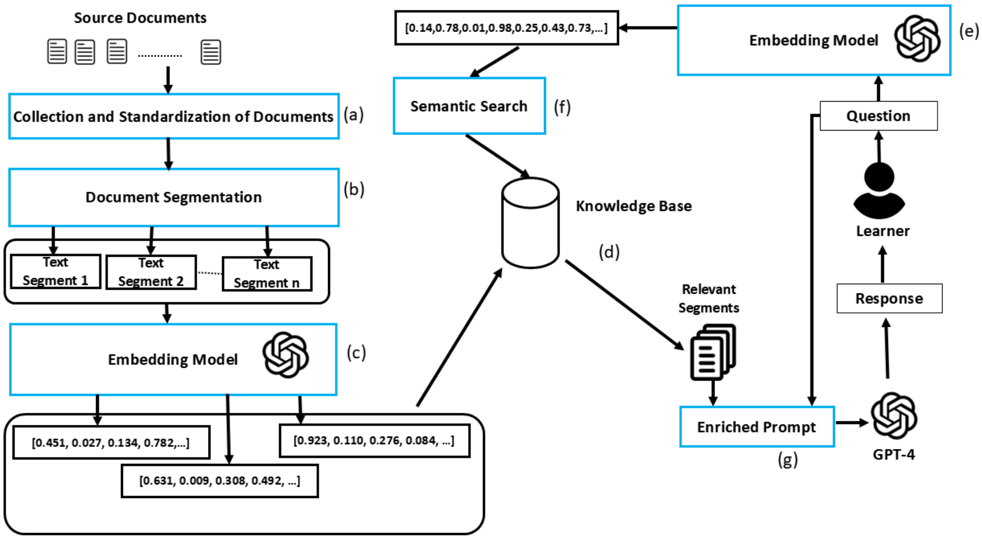

Figure 1: Architecture of the GPT-RAG conversational agent.

university in Senegal. Initially, there were 42 students in total, but 17 students did not complete the experiment for personal reasons. Consequently, the final number of research participants was 25. These participants were randomly divided into a control group (CG) (n=12; four females and eight males) and an experimental group (EG) (n=13; five females and eight males), with participants' ages ranging from 19 to 23.

4.2 Research Procedures

At the beginning of the study, students from both the CG and EG completed a pre-test to assess their understanding of artificial intelligence concepts. The experimental group watched a short tutorial on the conversational agent before using it in Module 1 of the AI MOOC. In contrast, the control group completed the same module without access to the conversational agent.

All participants worked individually and autonomously at their own pace, with three days to complete the task. To ensure timely completion, email reminders were sent on the second day to those who had not yet finished.

At the end of the experiment, all participants took a post-test to evaluate whether the chatbot significantly enhanced their knowledge acquisition. Additionally, participants in the experimental group completed a System Usability Scale (SUS) questionnaire to assess the chatbot's usability.

The experimental procedure is illustrated in Figure 2, providing a simplified draft of the key steps in the study. This figure highlights the sequence of activities, including pre-tests, post-tests, and the usability questionnaire conducted with the experimental group.

4.3 Research Instruments

The study employed various instruments to assess participants' knowledge acquisition and chatbot usability. To evaluate learners' understanding of AI in this MOOC, both groups completed a pre-test before the experiment and a post-test after Module 1 to measure knowledge acquisition. The tests included single-choice and short-answer questions, covering the same concepts to ensure consistency. The results helped address RQ1.

The System Usability Scale (SUS) (Brooke, 1996) was chosen for its simplicity, shortness, and reliability, even with a small sample size (Tullis and Stetson, 2004). The SUS consists of 10 statements, each rated on a 5-point Likert scale from "Strongly Disagree" (1 point) to "Strongly Agree" (5 points), producing a single usability score between 0 and 100. Higher scores indicate better usability. Odd-numbered statements reflect positive attitudes, while even-numbered statements reflect negative perceptions of the system. Responses to the SUS questionnaire were collected from 13 learners in the experimental group, who interacted with the conversational agent. This data provided insights to answer RQ2.

5 RESULTS

This section presents the quantitative analysis of the chatbot's impact on knowledge acquisition and us-

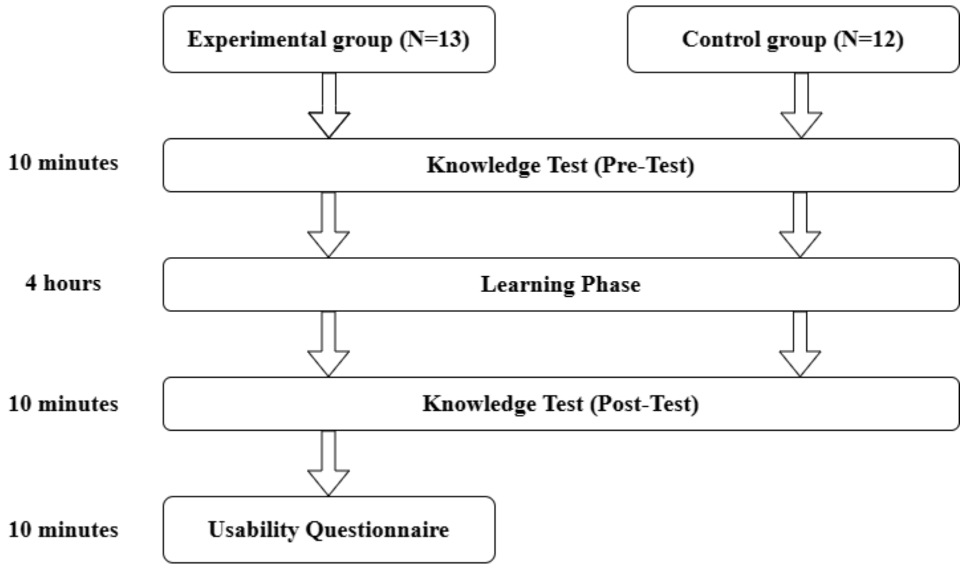

Figure 2: Experimental procedure.

ability. Knowledge acquisition was measured through pre- and post-tests, while the chatbot's usability was evaluated using the SUS.

5.1 Knowledge Gain Results

To evaluate knowledge acquisition in both the control and experimental groups, pre- and post-test assessments were conducted. The results, illustrated in Figure 3, show the percentage of knowledge gained by both groups. Initially, their average pre-test scores were similar (72%), indicating comparable prior knowledge levels.

After the learning activity, the experimental group, which used the chatbot, showed a 17% increase in knowledge gain, while the control group, without the chatbot, demonstrated a 10% gain. These results indicate that the chatbot had a positive effect on knowledge acquisition.

Statistical analysis confirmed these findings. Both groups showed improvement in their post-test scores, but the experimental group exhibited a more substantial increase. The statistical analyses of pre-test scores confirm that the control and experimental groups follow a normal distribution (Shapiro-Wilk test, $p > 0.05$) and have homogeneous variances (Levene's test, $p > 0.05$). These conditions allow for the application of a Student's t-test, which is appropriate for comparing the means of two independent groups when distributions are normal and variances are equivalent.

The t-test revealed no significant difference between the pre-test scores of the two groups ($p = 0.99 > 0.05$), indicating that both groups had similar levels of knowledge before the experiment (Table 1a).

However, a significant difference was observed in the post-test scores ($p = 0.017 < 0.05$), indicating that the conversational agent enhanced knowledge acquisition (Table 1b). The effect size was large ($d = 1.02$), indicating a substantial difference between the two groups.

5.2 SUS Results

A total of 13 responses were collected from the SUS questionnaire. Table 2 presents the detailed results for each questionnaire item, including the mean, median, and standard deviations for the responses.

Based on Brooke (1996), the overall SUS score is calculated by first adjusting the scores for both odd- and even-numbered questions. For the odd-numbered questions (questions 1, 3, 5, 7, and 9), 1 is subtracted from each score, and the resulting values are summed to compute the variable X. Similarly, for the even-numbered questions (questions 2, 4, 6, 8, and 10), each score is subtracted from 5, and these adjusted values are summed to compute the variable Y. The final SUS score is obtained by adding X and Y together and then multiplying the sum by 2.5, yielding a score that ranges from 0 to 100.

For our chatbot, the final SUS score was calculated as 80.4, indicating a high level of usability. SUS scores among learners ranged from 52.5 to 95 out of 100. Half of the users scored between 75 and 85, with a median score of 82.5.

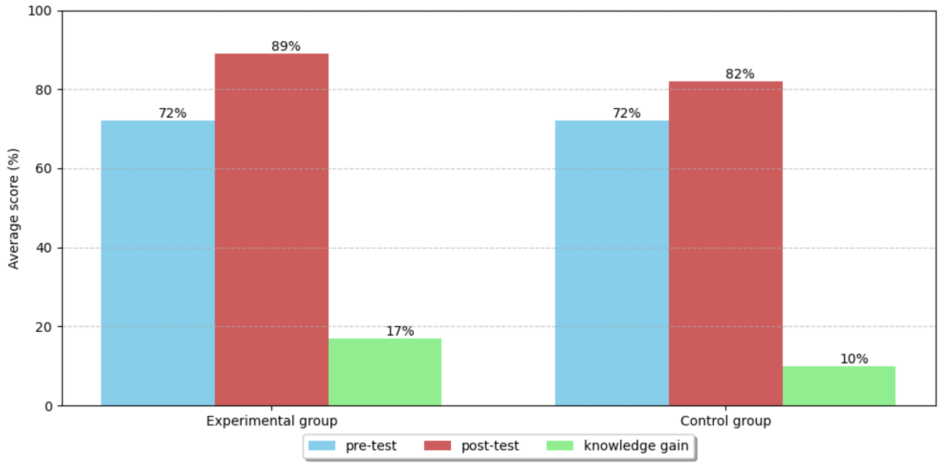

Figure 3: Average Pre- and Post-Test Scores for the EG and CG for Module 1 of the MOOC.

Table 1: Analysis of knowledge acquisition in Pre-test and Post-test.

(a) Pre-test

Group	N	Mean	Standard deviations	Median	P-value
Control	12	7.2	2.19	7.5	0.99
Experimental	13	7.2	1.9	8	

(b) Post-test

Group	N	Mean	Standard deviations	Median	P-value
Control	12	8.2	1.14	8	0.017
Experimental	13	8.9	1.03	9	

6 DISCUSSION

The findings suggest that the GPT-4-based chatbot enhanced with RAG improved knowledge acquisition. This improvement can be explained by the chatbot's ability to provide contextually relevant support in real time. By retrieving information from external sources, RAG reinforced the chatbot's generative capabilities, aiming to provide responses that were both accurate and adapted to learners' needs. This enhanced response quality likely helped clarify difficult concepts, contributing to the observed increase in knowledge gain.

Our results align with Slade et al. (2024), who evaluated a RAG-based tutoring system for writing assignments in an introductory psychology course. Their findings show that students using the system scored significantly higher on a post-test, suggesting improved knowledge retention. Similarly, Ko et al. (2024) investigated the integration of RAG with LLMs to enhance students' understanding and application of complex programming concepts. Their results indicate that learners using RAG achieved better results in solving unfamiliar problems, suggesting improved knowledge transfer and deeper conceptual understanding.

To address the second research question on chatbot usability, we used the SUS questionnaire. The SUS score obtained for our conversational agent is 80.4. According to Bangor et al. (2009), this corresponds to a "B" grade on the SUS rating scale. In terms of acceptability, the chatbot is classified as "Acceptable", and in adjective ratings, it falls under the "Good" category (see Figure 4). These results indicate that the chatbot is well received by learners and has strong potential to enhance user experience in educational settings.

This work is part of a paradigm change related to generative AI, marked by an increased use of conversational agents in learning, particularly in asynchronous distance learning contexts. These environments require a high degree of autonomy from learners, and conversational agents could represent a significant advancement in pedagogical support.

Table 2: SUS questionnaire and statistics for each item.

Question	Statement	Mean	Median	Standard deviations
1	I think that I would like to use this conversational agent.	4.46	4	0.50
2	I found the conversational agent unnecessarily complex.	1.85	2	0.86
3	I thought the conversational agent was easy to use.	4.54	5	0.63
4	I think that I would need the support of a technical person to be able to use this conversational agent.	1.15	1	0.36
5	I found the various functions in this conversational agent were well integrated.	3.85	4	0.77
6	I thought there was too much inconsistency in this conversational agent.	1.54	1	0.84
7	I would imagine that most people would learn to use this conversational agent very quickly.	4.38	4	0.62
8	I found the conversational agent very cumbersome to use.	2.15	2	1.10
9	I felt very confident using the conversational agent.	4.31	4	0.72
10	I needed to learn a lot of things before I could get going with this conversational agent.	2.69	3	1.43

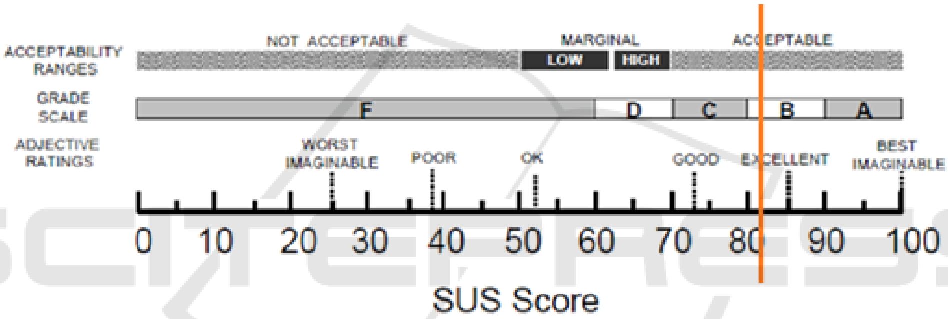

Figure 4: SUS Bangor Scale (Bangor et al., 2009) and SUS score for conversational agent (Mean Value).

In this context, conversational agents function as learning companions, as envisioned by Chan and Baskin (1988), providing adaptive support based on learners' needs. They leverage their superior knowledge while remaining susceptible to occasional errors. Rather than replacing teachers or human experts, they function as interactive learning companions, particularly in contexts with limited instructional support.

This companion role is especially crucial in non-credit distance courses, such as MOOCs, where learners must navigate content independently. By delivering contextualized and tailored responses, RAG-enhanced conversational agents help sustain learner engagement, mitigating the risk of dropout in online education.

7 CONCLUSIONS

This study suggests that a GPT-4-powered conversational agent enhanced with RAG improves knowledge acquisition in MOOCs. By delivering real-time, contextually relevant support, the chatbot appears to support learners' understanding of course content and promote a more engaging learning experience. The results indicate a statistically significant improvement in knowledge gain, along with positive learner perceptions of usability, reinforcing the potential of RAG-enhanced AI in online education.

Despite promising results, this study has limitations, notably a small, single-institution sample that restricts generalizability, particularly in the context of MOOCs, where large-scale dynamics are essential. Additionally, the short study duration limited the ability to assess long-term learning effects. Future research should incorporate a larger and more diverse participant group, extend the study period, and further evaluate the chatbot's effectiveness in large-scale MOOC environments.

Future work will focus on designing an empathetic conversational agent based on LLMs and RAG, capable of detecting learners' emotions in real time and adapting its interactions accordingly. By tailoring responses to learners' emotions and needs, the agent could enhance engagement, persistence, and learning outcomes. Further development will refine its emotion recognition capabilities to optimize interactions and create a more adaptive and enriching educational experience.

REFERENCES

Abdelghani, R., Wang, Y., Yuan, X., Wang, T., Sauzéon, H., and Oudeyer, P. (2022). Gpt-3-driven pedagogical agents for training children's curious question-asking skills. arxiv. *preprint arXiv*, 2211.

Adeshola, I. and Adepoju, A. P. (2024). The opportunities and challenges of chatgpt in education. *Interactive Learning Environments*, 32(10):6159–6172.

Bangor, A., Kortum, P., and Miller, J. (2009). Determining what individual sus scores mean: Adding an adjective rating scale. *Journal of usability studies*, 4(3):114–123.

Brooke, J. (1996). Sus: A quick and dirty usability scale. *Usability Evaluation in Industry*.

Chan, T.-W. and Baskin, A. B. (1988). Studying with the prince: The computer as a learning companion. In *Proceedings of the International Conference on Intelligent Tutoring Systems*, volume 194200.

Chiu, T. K. (2024). Future research recommendations for transforming higher education with generative ai. *Computers and Education: Artificial Intelligence*, 6:100197.

Chiu, T. K. and Hew, T. K. (2018). Factors influencing peer learning and performance in mooc asynchronous online discussion forum. *Australasian Journal of Educational Technology*, 34(4).

Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., and Wang, H. (2023). Retrieval-augmented generation for large language models: A survey. *arXiv preprint arXiv:2312.10997*.

Hone, K. S. and El Said, G. R. (2016). Exploring the factors affecting mooc retention: A survey study. *Computers & Education*, 98:157–168.

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., and Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12):1–38.

Ko, H.-T., Liu, Y.-K., Tsai, Y.-C., and Suen, S. (2024). Enhancing python learning through retrieval-augmented generation: A theoretical and applied innovation in generative ai education. In *International Conference on Innovative Technologies and Learning*, pages 164–173. Springer.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., et al. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 33:9459–9474.

Liu, R., Zenke, C., Liu, C., Holmes, A., Thornton, P., and Malan, D. J. (2024). Teaching cs50 with ai: leveraging generative artificial intelligence in computer science education. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1*, pages 750–756.

Mariani, M. M., Hashemi, N., and Wirtz, J. (2023). Artificial intelligence empowered conversational agents: A systematic literature review and research agenda. *Journal of Business Research*, 161:113838.

Miladi, F., Psyché, V., and Lemire, D. (2024). Leveraging gpt-4 for accuracy in education: A comparative study on retrieval-augmented generation in moocs. In *International Conference on Artificial Intelligence in Education*, pages 427–434. Springer.

Neelakantan, A., Xu, T., Puri, R., Radford, A., Han, J. M., Tworek, J., Yuan, Q., Tezak, N., Kim, J. W., Hallacy, C., et al. (2022). Text and code embeddings by contrastive pre-training. *arXiv preprint arXiv:2201.10005*.

Psyché, V. (2020). Clom-Motsia: MOOC sur l'intelligence artificielle. https://clom-motsia.teluq.ca/, last accessed Jan 17 2024.

Shuster, K., Poff, S., Chen, M., Kiela, D., and Weston, J. (2021). Retrieval augmentation reduces hallucination in conversation. *arXiv preprint arXiv:2104.07567*.

Slade, J. J., Hyk, A., and Gurung, R. A. (2024). Transforming learning: Assessing the efficacy of a retrieval-augmented generation system as a tutor for introductory psychology. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 68, pages 1827–1830. SAGE Publications Sage CA: Los Angeles, CA.

Taneja, K., Maiti, P., Kakar, S., Guruprasad, P., Rao, S., and Goel, A. K. (2024). Jill watson: A virtual teaching assistant powered by chatgpt. In *International Conference on Artificial Intelligence in Education*, pages 324–337. Springer.

Tullis, T. S. and Stetson, J. N. (2004). A comparison of questionnaires for assessing website usability. In *Usability professional association conference*, volume 1, pages 1–12. Minneapolis, USA.

Vijaymeena, M. and Kavitha, K. (2016). A survey on similarity measures in text mining. *Machine Learning and Applications: An International Journal*, 3(2):19–28.

Wang, K., Ramos, J., and Lawrence, R. (2023). Chated: a chatbot leveraging chatgpt for an enhanced learning experience in higher education. *arXiv preprint arXiv:2401.00052*.

Xie, Z., Wu, X., and Xie, Y. (2024). Can interaction with generative artificial intelligence enhance learning autonomy? a longitudinal study from comparative perspectives of virtual companionship and knowledge acquisition preferences. *Journal of Computer Assisted Learning*.

Yan, L., Zhao, L., Echeverria, V., Jin, Y., Alfredo, R., Li, X., Gaševi'c, D., and Martinez-Maldonado, R. (2024). Vizchat: enhancing learning analytics dashboards with contextualised explanations using multimodal generative ai chatbots. In *International Conference on Artificial Intelligence in Education*, pages 180–193. Springer.

Yin, S., Shang, Q., Wang, H., and Che, B. (2019). The analysis and early warning of student loss in mooc course. In *Proceedings of the ACM Turing Celebration Conference-China*, pages 1–6.