Jurnal Inovasi Teknologi Pendidikan
Volume 13, No. 1, March 2026 (57-69)
Online: http://journal.uny.ac.id/index.php/jitp

Enhancing student academic services through AI-driven virtual assistants using the RAG method at Universitas Terbuka

Adhi Susilo[1] *, Athiefah Fauziyyah[1], Ariyanti Hartari[1], Dian Nurdiana[1], Eko Yuliastuti Endah Sulistyawati[1], Abdul Rizal Adompo[1], Isfarudi[1], Alaa Abdelkader Abdelmonem Abdelkader[2]
[1] Universitas Terbuka, Indonesia.
[2] CLT Cairo University, Egypt.
* Corresponding Author. E-mail: adhi@ecampus.ut.ac.id

ARTICLE INFO

Article History
Received:
28 February 2025;
Revised:
10 June 2025;
Accepted:
10 June 2025;
Available online:
31 March 2026.

Keywords
Chatbot; Large language model; RAG; Virtual assistant

ABSTRACT

The Indonesian Open University currently operates an information service called Hallo UT, which mainly provides academic and administrative information. To improve this service, this study develops a virtual assistant chatbot capable of delivering autonomous 24-hour customer support using a Retrieval-Augmented Generation (RAG) approach. RAG combines information retrieval techniques with large language model capabilities to generate accurate and contextually relevant responses. Data were collected from academic manuals containing frequently asked questions and questionnaires distributed to 76 students. The chatbot was evaluated based on accuracy, response time, and user satisfaction. Results show that the system achieved an average accuracy rate of 92% with an average response time of 5 seconds. In addition, 62% of students responded positively to the chatbot's functionality. These findings demonstrate the chatbot's potential to improve student engagement, reduce administrative workload, and enhance the overall learning experience. Future research should involve larger samples, multilingual support, and broader system integration.

How to cite:
Susilo, A., et al., (2026). Enhancing student academic services through AI-driven virtual assistants using the RAG method at Universitas Terbuka. *Jurnal Inovasi Teknologi Pendidikan, 13*(1), 57-69. https://doi.org/10.21831/jitp.v13i1.83632

INTRODUCTION

In the ever-evolving landscape of higher education, universities and colleges continuously seek new ways to enhance academic services and student support. One promising approach is the use of AI-powered virtual assistants, which can provide personalised, adaptive support to students throughout their learning process (Berkat, 2024).

AI has had a significant impact on e-learning, offering viable solutions to address the limitations of conventional e-learning approaches. AI-powered digital assistants can tailor learning by adjusting content, pacing, and feedback according to the specific needs of each student (Kashive et al., 2021). Moreover, adaptive assessment, another application of AI in e-learning, can track student progress and provide targeted feedback to support learning (Rasheed et al., 2023). Advanced implementations, such as smart tutoring systems, offer personalised guidance and support, improving both learning outcomes and student engagement (Castro et al., 2024).

As Indonesia's leading distance learning institution, Universitas Terbuka (UT) continues to improve the quality of its academic services. One of the innovative efforts underway is the development of an AI-based virtual assistant to support real-time interaction and information needs. This study specifically examines the application of the retrieval-augmented generation (RAG) model to develop an intelligent chatbot for UT. Supported by RAG, this chatbot is designed to handle a wide range of student inquiries, from course-specific information to general academic support, to increase efficiency and accessibility in a distance learning environment (Vázquez et al., 2021).

A chatbot is an AI-powered conversational service that simulates human interaction. This technology can understand and process user requests, delivering quick, relevant responses (Dongbo et al., 2023). It operates by scanning keywords from user input and matching them to predefined patterns or keyword sets, generating automatic responses based on the best match (Dongbo et al., 2023).

To date, Universitas Terbuka has not implemented a chatbot service. The university lacks a 24-hour interactive communication platform. Customer service at UT's Centre for Information and Communication Technology (PTIK) is not available around the clock, even though students often need information outside regular working hours. Chatbot technology presents a promising solution to this issue, offering the potential to serve as a virtual assistant, replacing conventional customer service by providing timely information and communication support (Santos et al., 2022).

Previous studies have highlighted the potential of chatbots in education. For instance, a study explored the use of chatbots to respond to administrative queries at a private university (Ramakrishnan et al., 2024), while another study investigated user satisfaction with mobile-based university chatbots (Nasa et al., 2023). Additionally, research emphasised how chatbot systems can enhance student learning experiences through responsive two-way interaction (Baha et al., 2024). Another study found that implementing chatbots in higher education institutions could reduce the workload of information service staff by up to 40% (Popescu et al., 2023). However, most of these studies focus on traditional or rule-based chatbot systems, with limited attention to the context of large-scale distance education, such as that of Universitas Terbuka. This presents a research gap: how to design and implement an AI-powered chatbot using a modern RAG approach to support real-time academic services in distance learning settings.

Furthermore, implementing AI in academic services is not without challenges. One major issue is the availability and quality of internal data needed to train AI models that can accurately respond to student queries (Su & Yang 2023). Based on internal observations, approximately 40% of student inquiries to customer service are repetitive and administrative, such as exam schedules, re-registration procedures, or assignment submission guidelines. Yet, they are not systematically documented in a database that an AI system could readily utilise.

This study aims to implement a retrieval augmented generation-based chatbot as a virtual assistant at Universitas Terbuka to improve the effectiveness of real-time academic information services for students. The key contribution of this research is the development of an AI-based chatbot implementation model suited to distance education contexts, along with a comprehensive analysis of the technical and non-technical challenges encountered during its development. The findings are expected to serve as a reference for other higher education institutions seeking to adopt AI technologies to deliver more responsive and inclusive academic services.

METHOD

The development of the AI-based virtual assistant at the Open University of Indonesia followed a multi-step process: (1) Knowledge base curation, the university's academic and administrative data, including course materials, syllabi, and student support resources, were compiled to create a comprehensive knowledge base for the virtual assistant (Sugianto et al., 2021). (2) Language processing, a natural language processing module was developed to understand and interpret student queries, leveraging techniques such as intent recognition and entity extraction (Sajja et al., 2023). (3) Natural language processing, a natural language processing module was developed to understand and interpret student queries, enabling the virtual assistant to provide relevant and personalised responses (Tu et al., 2023). (4) Retrieval augmented generation, the retrieval augmented generation method was implemented to generate contextually appropriate responses by combining language model-based generation with information retrieval from the knowledge base (Jiang et al., 2023). (5) Adaptive assessment, the virtual assistant's capabilities were extended to include adaptive assessment, allowing the system to evaluate student progress and provide targeted feedback and support continuously (Kadaruddin, 2023). (6) Integration with learning management system, the AI-based virtual assistant was seamlessly integrated with the university's learning management system, providing students with a unified and accessible academic support platform (Latif et al., 2024; Lee et al., 2024).

In addition, some steps have been taken to implement Retrieval-Augmented Generation (RAG) in this research: (1) Data collection, gathering relevant documents, articles, FAQs, and website knowledge bases. Focus on comprehensiveness, ensuring the data covers a wide range of topics a chatbot might encounter. (2). Data transformation, implement an Extract-Transform-Load (ETL) process to combine data from various sources into a single, unified format. This step involves cleaning, structuring, and standardising the data for consistency. (3) Data vectorisation, load the preprocessed data from step 2 into ChromaDB, a vector database. ChromaDB stores data in a format that enables efficient similarity-based searching. This involves creating vector representations of the data points, enabling faster retrieval of relevant information during user queries. (4) Query and retrieval, this stage combines retrieval and generation, the core of RAG. Retrieval: When a user asks a question, the system formulates a query based on the user's input. This query is then used to search ChromaDB for the most relevant documents. The search leverages the vector representations created earlier to find documents with high semantic similarity to the user's query. (5) Augmentation: the retrieved documents from ChromaDB are then combined with the original user query. This creates a richer context, providing additional information for the large language model. Essentially, we provide the LLM with both the user's question and related information from the knowledge base. (6) Response generation with LLM, the augmented query, containing both the user input and relevant retrieved documents, is sent to OpenAI's large language model (LLM). The LLM leverages its capabilities and the provided context to generate a response. By using the retrieved information, the LLM can produce more informative and accurate answers that are grounded in factual knowledge. This is a simplified overview of implementing RAG. There are additional considerations, such as choosing the appropriate data sources, fine-tuning the retrieval process, and potentially using different vectorisation techniques or LLMs. The implementation of Retrieval-Augmented Generation (RAG) in this study is shown in Figure 1 below.

Figure 1. AI Virtual Assistant Development Process

Testing the Accuracy Rate of Chatbot

The testing accuracy rate can be calculated using the accuracy formula as presented in Formula 1 below.

$$Accuracy = \frac{Total\ Right\ Answers}{Total\ Questions} 100\% \quad (1)$$

User-Acceptance Test

A User Acceptance Test is a test conducted to verify that the software built to meet existing needs is acceptable. The purpose of UAT is to identify the determinants of general computer acceptance that explain user behaviour across various computing technologies. In this study, the UAT was conducted with 76 student respondents from Universitas Terbuka (UT) to assess the effectiveness and acceptance of the AI-based virtual assistant system. The Questionnaire Instrument is shown in Table 1.

Table 1. Questionnaire Instrument

No.	Questions
1	Do you agree that the chatbot's response time is fast enough?
2	Do you agree that the responses provided by the chatbot are relevant to the questions you asked?
3	Do you agree that the chatbot introduces itself well enough for users to understand how to use it?
4	Do you agree that the chatbot functions properly?
5	Do you agree? How does the chatbot help you obtain academic-related information?
6	Do you agree that this chatbot makes it easier and faster for you to obtain information related to academic and administrative services at Universitas Terbuka?
7	Do you agree that the chatbot can serve as your primary virtual assistant for obtaining academic and administrative information at Universitas Terbuka?
8	Do you agree that the chatbot helps you get information related to academic and administrative matters?

RESULTS AND DISCUSSION

Results

The retrieval-augmented generation method has proven effective at generating contextually appropriate responses, drawing on the knowledge base to provide comprehensive and personalised support to students.

The virtual assistant has responded to many student queries with high accuracy and relevance, as evidenced by positive student feedback and improved satisfaction (Chheang et al., 2024). The adaptive assessment capabilities of the virtual assistant have also enabled the university to continuously monitor student progress and provide targeted support, thereby improving learning outcomes.

The retrieval-augmented generation method has enabled the virtual assistant to provide comprehensive, contextually appropriate responses, drawing on the rich knowledge base curated by the university.

Chatbot System Architecture

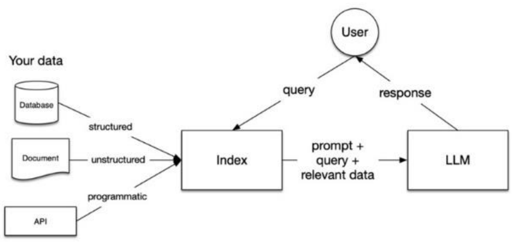

Figure 2. Chatbot System Architecture

is the architecture of the chatbot system. At this stage, text preprocessing is performed, including word weighting, document similarity calculation, and ranking by the highest weight value. When the user sends a message (query), the system processes the data by going through the text preprocessing stage, then calculates the data using the TF-IDF and VSM methods, and computes similarity using cosine similarity.

There are three main approaches to AI technology, namely fine-tuning, embedding, and Retrieval-Augmented Generation (RAG). Fine-tuning in AI involves adjusting pre-trained models to specific tasks, and slight variations in the fine-tuning process can significantly impact performance (Goyal et al., 2023; Zhang & Hu, 2021). Embedding refers to representing data in a lower-dimensional space, enhancing semantic similarity tasks without fine-tuning, as seen in music similarity research (Zhang et al., 2021). On the other hand, the Retrieval Augment Generation (RAG) approach focuses on end-to-end fine-tuning of the RAG architecture, surpassing the original model's performance in question answering tasks (Siriwardhana et al, 2021). While fine-tuning emphasises task-specific adjustments, embedding techniques aim to improve model performance without extensive retraining, and RAG highlights the benefits of end-to-end fine-tuning for specific architectures, showcasing the diverse strategies in AI technology. The AI Chatbot Based on the RAG Method as a Virtual Assistant is shown in Figure 3 below.

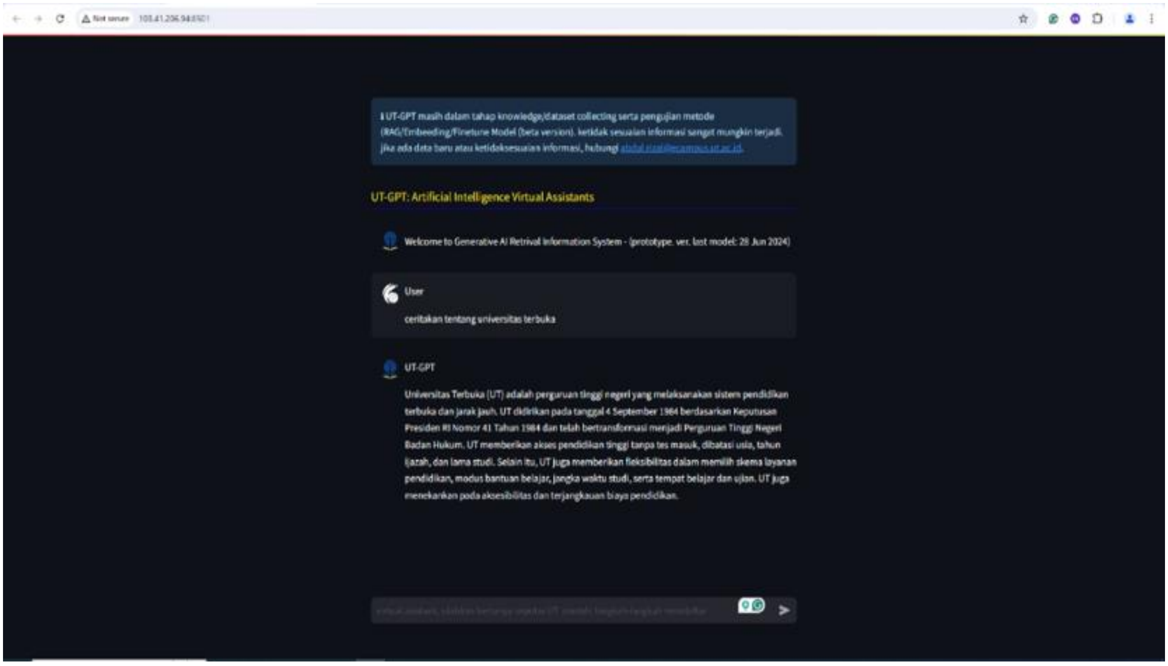

Figure 3. Chatbot AI-Based on RAG Method as a Virtual Assistant in UT

Testing Accuracy Rate

Testing Accuracy Rate is conducted using a dataset of 50 questions. Based on the test results, the system correctly answered 46 out of 50 questions, achieving an accuracy of 92%. These results indicate that the chatbot system has a relatively high level of accuracy in classifying and responding to user queries. However, several obstacles contribute to the remaining inaccuracies. One of the main issues is the presence of essential words across multiple sentence classes, which makes it difficult for the system to classify questions correctly. In addition, the use of non-standard Indonesian abbreviations, such as “mk” (mata kuliah/course) and “PS” (Program Studi/department), cannot be properly recognised by the system. Furthermore, informal and non-standard words such as “tak” (*tidak*/no) and “y” (*ya*/yes) are not processed effectively after tokenisation. These limitations reduce the system’s ability to achieve maximum accuracy in question classification.

User-Acceptance Test

Testing the chatbot application on 76 students. Testing respondents in four categories using chatbots, including ability, consistency, responsibility, and performance. Based on a questionnaire distributed to users, the chatbot was found to work well.

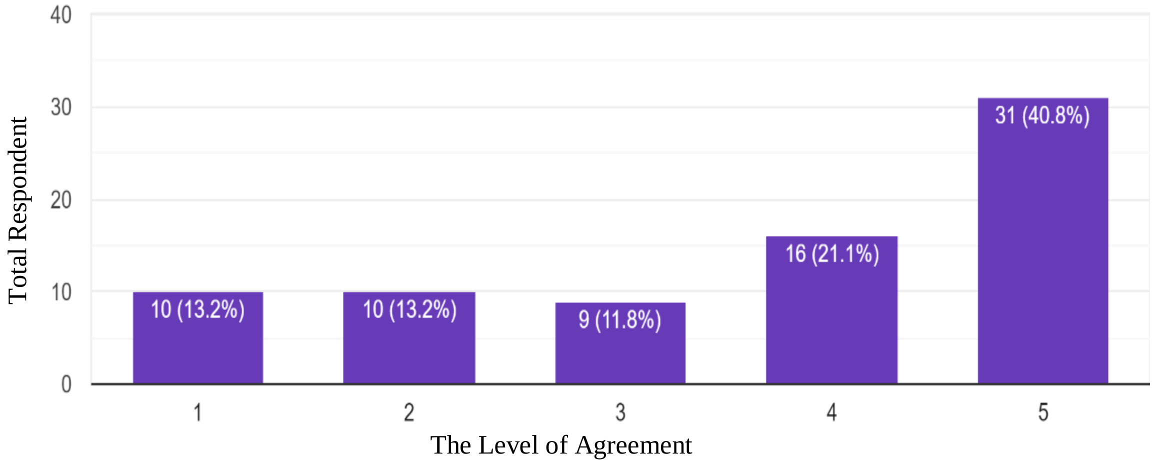

Figure 4. User Opinion on Chatbot Performance

Notes:

a. Strongly Disagree: Indicates the lowest level of agreement or strongest negative perception.
b. Disagree: Reflects disagreement with the statement but with less intensity than "Strongly Disagree".
c. Neutral: Represents the midpoint, indicating neither agreement nor disagreement, or ambivalence.
d. Agree: Implies agreement with the statement, though less intense than "Strongly Agree".
e. Strongly Agree: Represents the highest level of agreement or strongest positive perception.

Users agree that the chatbot can be a top priority as a virtual assistant to get information about Open University academics and administration.

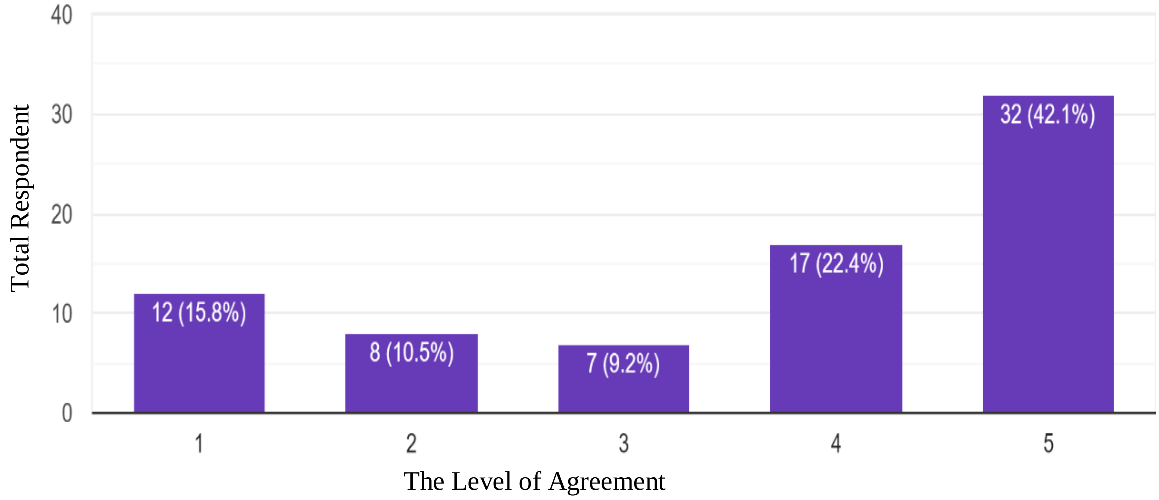

Figure 5. User Opinion on Chatbot as a Virtual Assistant

From these data, in Figures 4 and 5, it is evident that more than 60% of users state that using chatbot technology applications is better in terms of ability, consistency, responsiveness, and performance.

Discussion

Finetuning enhances AI model performance by adapting the weights of specific layers or groups of layers in pre-trained models to tackle domain-specific tasks effectively (Barakat & Huang, 2023). Traditional methods that focus on task-specific classifier weights or re-optimising all layers may lead to overfitting with limited data or fail to address the mismatch between pre-trained models and new-task data (Barakat & Huang, 2023). To address these challenges, novel block-wise optimisation mechanisms have been proposed, allowing for the adaptation of groups of layers in pre-trained models through various strategies like layer-wise adaptation, joint adaptation of top-ranked layers, block-based segmentation, and sliding window grouping, resulting in improved performance compared to baseline methods and layer-wise approaches (Barakat & Huang, 2023). Additionally, reinforcement learning-based finetuning methods have shown promise for enhancing AI systems' understanding of document images, especially in scenarios with limited training data, by jointly optimising combined reward functions alongside traditional losses (Nguyen et al., 2022).

Recent embedding techniques play a crucial role in enhancing AI model accuracy by optimising computation, reducing computational costs, and improving generalisation. Techniques such as network pruning, low-precision quantisation, and dynamic inference help compress models, thereby minimising computational and storage intensity (Liu, 2021). Additionally, approaches that replace data augmentation in the raw input space with an approximate augmentation in the embedding space significantly reduce computational cost while maintaining model accuracy (Abrishami et al., 2020). Moreover, using specific embedding strategies, such as GraNet layers in Graph Neural Networks, enables neighbourhood aggregation and the inheritance of weights from pre-trained models, leading to improved accuracy compared to traditional methods (Purchase et al., 2022). Furthermore, supervised document embedding designed for hierarchical text categorisation, trained on both words and labels, enhances embedding vectors by leveraging class hierarchy information, resulting in superior performance in text categorisation tasks (Saetia & Vateekul, 2018).

Recent advancements in embedding techniques have significantly impacted AI model performance. Techniques such as network pruning, low-precision quantisation, and dynamic inference have been proposed to compress models, reducing computational and storage requirements (Liu, 2021). Additionally, approaches such as transferring data augmentation from the raw input space to the embedding space have been introduced to decrease computational costs while maintaining model accuracy (Abrishami et al., 2020). Furthermore, optimising the deployment of embeddings using frameworks such as Hetero-Rec for recommendation models has shown substantial improvements in reducing inference latency by caching frequently accessed embeddings in faster memory (Mahajan et al., 2022). Tools like Emblaze have been developed to aid in comparing embedding spaces, enabling model builders to choose optimal representations and identify flaws for improved model performance (Sivaraman et al., 2022). These advancements enhance the efficiency and effectiveness of AI models.

Retrieval Augment Generation (RAG) is a cutting-edge technique in Open-Domain Question Answering (ODQA) that combines a retriever and a generator to enhance AI model accuracy. RAG has traditionally been trained on a Wikipedia-based knowledge base, limiting its adaptability to specialised domains such as healthcare and news (Siriwardhana et al., 2023). RAG-end2end, an extension of RAG, enables joint training of the retriever and generator components, facilitating domain adaptation by updating all knowledge base components during training and injecting domain-specific knowledge through an auxiliary training signal (Siriwardhana et al., 2023). Additionally, RAG has been successfully applied in automating radiology report writing, leveraging multimodal embeddings for retrieval and generative models for report generation, resulting in improved clinical metrics and the ability to tailor report content to specific clinical settings (Ranjit et al., 2023).

The rollout of AI-powered virtual assistants in higher education, such as those at the Open University of Indonesia, has had a significant impact on student academic services. These tools help in many ways. They provide quick, correct answers to student questions using the retrieval-augmented generation method. They also boost learning outcomes through adaptive assessment, which helps spot and address learning gaps. Linking the virtual assistant with the university's learning management system also made a single, easy-to-use platform for students who need academic help

(Sajja et al. 2023). Using the AI-powered virtual assistant at the Open University of Indonesia has shown clear gains in student academic services (Muhammad & Sudianto, 2023). The virtual assistant can answer many student questions, from course-related queries to general academic help, with high accuracy and relevance (Rios et al., 2023).

The retrieval-augmented generation method has been shown to produce responses that fit the context. It leverages the extensive information in the knowledge base to provide thorough, personalised help to students (Abu & Alotaibi, 2024). The virtual assistant has answered many different student questions, and students have given good feedback and are happier with the service (Chheang et al., 2024). The virtual assistant can also adapt its assessments. This allows the university to track how students are doing and provide help when they need it. As a result, students are learning better (Dogan et al., 2023).

The virtual assistant uses a method called retrieval augmented generation. This helps it provide full and fitting answers by drawing on the university's extensive body of knowledge (Wang et al., 2023). The virtual assistant's integration with the university's learning management system has created an easy-to-use, accessible academic support platform for students. This has boosted the impact of the AI-powered solution (Sajja et al., 2023). The AI-based virtual assistant rollout has succeeded, but the university still faces ongoing issues and considerations (Hajipour et al., 2023). These include ensuring data remains private and secure, reducing bias in algorithms, and continually improving the virtual assistant through machine learning and user feedback (Rios et al., 2023).

The retrieval-augmented generation method enhances the virtual assistant's ability to provide students with quick, correct answers to questions across many academic topics, from course-specific questions to general academic help (Sajja et al., 2023). The retrieval-augmented generation method combines the strengths of large language models with a vast knowledge base. This allows the virtual assistant to access and use information from various sources, such as course materials, syllabi, and student support resources (Gill et al., 2024).

When a student asks a question, the virtual assistant first uses natural language processing to understand the student's meaning and identify keywords (Hajipour et al., 2023). Then, it uses the retrieval-augmented generation method to search its knowledge base for the most accurate and up-to-date information (Zhang et al., 2024). It produces a full-fitting answer by combining its ability to generate language with the information it found (Walker et al., 2023). This method enables the virtual assistant to address many school-related issues, from specific questions about class material to general inquiries about school rules or help services, ensuring students receive quick, accurate information (Chheang et al., 2024).

The virtual assistant's ability to adapt its assessments has also shown its worth, enabling the university to identify learning gaps and offer targeted support to students, thereby improving learning outcomes (Dogan et al., 2023). Also, combining the AI-based virtual assistant with the university's learning management system has created an easy-to-use platform for students to access academic help, thereby improving their overall learning experience (Sajja et al., 2023). Putting the AI-based virtual assistant into action hasn't been smooth sailing. The team had to tackle some tricky issues. They took a hard look at ethical concerns, such as keeping people's data safe and ensuring the system treats everyone equally (TonbuloĞLu, 2023). Moreover, chatbots' ability to understand the variety of informal language, slang, and idiomatic expressions used by students in daily interactions is limited (Shams et al., 2024). In the increasingly diverse context of higher education, students not only use formal language when communicating but also mix technical terms, everyday language, and even popular social media abbreviations. When chatbots fail to understand expressions such as "tight credit hours", "sudden academic leave", or "grades not yet input", the user experience can be disrupted, and trust in the technology may decline. Future chatbot development should integrate natural language processing models that are more adaptive to non-standard and context-specific language (Suryanto et al., 2023). Approaches such as continual learning and contextual fine-tuning based on real student conversational data can be employed to broaden the chatbot's understanding scope. Additionally, integration with corpora of informal student conversations or campus social media could be a strategic step to enrich the chatbot's semantic comprehension (Lin et al., 2025). By comparison, a study by Belda & Calvo (2022) shows that chatbots designed for informal educational settings and online communities perform better when trained using datasets of representative informal language. Therefore, in the future, chatbot design in campus environments should consider the linguistic and sociocultural aspects of students as primary users, ensuring that this technology is not only technically intelligent but also contextually relevant.

The launch of the AI virtual assistant at Universitas Terbuka Indonesia has shown promising results. The assistant can answer various student questions with high accuracy and relevance. The AI system architecture consists of a natural language processing (NLP) unit, an information retrieval unit, and an answer generation unit. This study's findings have significant potential to influence the design, implementation, and evaluation of AI-based Virtual Teaching Assistants in higher education. These findings could drive the creation of innovative learning tools to improve learning outcomes, student engagement, and satisfaction (Sajja et al., 2024). Moreover, more than 60% of users state that chatbot applications are superior in terms of ability, consistency, responsiveness, and performance compared to traditional service methods. This aligns with the findings of (Seraquive et al., 2024), which show that personalised and contextual chatbots can significantly enhance user satisfaction. The study by Seraquive et al., (2024) further supports this claim, where users felt that the interaction experience with chatbots was more consistent and responsive, especially in the context of academic and administrative support. These findings confirm the great potential of chatbot technology to improve the efficiency of higher education services.

CONCLUSION

The Open University of Indonesia's case study on using an AI virtual assistant shows how these smart systems can boost student academic services. The virtual assistant uses a method called retrieval-augmented generation to provide personalised, relevant answers. Its adaptive assessment features have helped students learn better. This research could shape how colleges design, roll out, and check AI-powered Virtual Teaching Assistants. It paves the way for new learning tools that can boost student learning, involvement, and happiness. We still need to study the long-term effects of AI virtual assistants on students' academic performance. We also need to consider the ethical implications of using this tech in education. For future research, it is recommended to expand the sample size and the diversity of respondents to gain more generalizable insights, incorporate multilingual support to accommodate a broader student population, and explore integrating the chatbot with other university systems to provide a more comprehensive academic service platform.

REFERENCES

Abrishami, M. S., Eshratifar, A. E., Eigen, D., Wang, Y., Nazarian, S., & Pedram, M. (2020). Efficient Training of deep convolutional neural networks by augmentation in embedding space. *arXiv*, 1-5. https://doi.org/10.48550/arXiv.2002.04776

Abu-Salih, B., & Alotaibi, S. (2024). A systematic literature review of knowledge graph construction and application in education. *Heliyon*, *10*(3), 1-23. https://doi.org/10.1016/j.heliyon.2024.e25383

Baha, T. A., El Hajji, M., Es-Saady, Y., & Fadili, H. (2024). The impact of educational chatbot on student learning experience. *Education and Information Technologies*, *29*(8), 10153-10176. https://doi.org/10.1007/s10639-023-12166-w

Barakat, B., & Huang, Q. (2023). Improving reliability of fine-tuning with block-wise optimisation. *arXiv*, *20*(20), 1-10. https://doi.org/10.48550/arXiv.2301.06133

Belda-Medina, J., & Calvo-Ferrer, J. R. (2022). Using chatbots as AI conversational partners in language learning. *Applied Sciences*, *12*(17), 1-16. https://doi.org/10.3390/app12178427

Berkat. (2024). Education transformation in Palangka Raya: Optimizing technology management through artificial intelligence towards quality improvement. *JTP-Jurnal Teknologi Pendidikan*, *26*(1), 131-140. https://doi.org/10.21009/jtp.v26i1.43402

Castro, G. P. B., Chiappe, A., Becerra Rodriguez, D. F., & Sepulveda, F. G. (2024). Harnessing AI for education 4.0: Drivers of personalized learning. *Electronic Journal of e-Learning*, *22*(5), 01-14. https://doi.org/10.34190/ejel.22.5.3467

Chheang, V., Sharmin, S., Márquez-Hernández, R., Patel, M., Rajasekaran, D., Caulfield, G., Kiafar, B., Li, J., Kullu, P., & Barmaki, R. L. (2024). Towards anatomy education with generative ai-based virtual assistants in immersive virtual reality environments. *arXiv*, 1-11. https://doi.org/10.48550/arXiv.2306.17278

Dogan, M. E., Goru Dogan, T., & Bozkurt, A. (2023). The use of artificial intelligence (AI) in online learning and distance education processes: A systematic review of empirical studies. *Applied Sciences*, *13*(5), 1-12. https://doi.org/10.3390/app13053056

Dongbo, M., Miniaoui, S., Fen, L., Althubiti, S. A., & Alsenani, T. R. (2023). Intelligent chatbot interaction system capable for sentimental analysis using hybrid machine learning algorithms. *Information Processing & Management*, *60*(5), 103440. https://doi.org/10.1016/j.ipm.2023.103440

Gill, S. S., Xu, M., Patros, P., Wu, H., Kaur, R., Kaur, K., Fuller, S., Singh, M., Arora, P., Parlikad, A. K., Stankovski, V., Abraham, A., Ghosh, S. K., Lutfiyya, H., Kanhere, S. S., Bahsoon, R., Rana, O., Dustdar, S., Sakellariou, R., Uhlig, S., & Buyya, R. (2024). Transformative effects of chatGPT on modern education: Emerging era of AI chatbots. *Internet of Things and Cyber-Physical Systems*, *4*, 19-23. https://doi.org/10.1016/j.iotcps.2023.06.002

Goyal, S., Kumar, A., Garg, S., Kolter, Z., & Raghunathan, A. (2023). Finetune like you pretrain: Improved finetuning of zero-shot vision models. *arXiv*, 1-19. https://doi.org/10.48550/arXiv.2212.00638

Hajipour, V., Hekmat, S., & Amini, M. (2023). A value-oriented artificial intelligence-as-a-service business plan using integrated tools and services. *Decision Analytics Journal*, *8*, 1-15. https://doi.org/10.1016/j.dajour.2023.100302

Jiang, Z., Xu, F., Gao, L., Sun, Z., Liu, Q., Dwivedi-Yu, J., Yang, Y., Callan, J., & Neubig, G. (2023). Active retrieval augmented generation. *arXiv*, 1-24. https://doi.org/10.48550/arXiv.2305.06983

Kadaruddin, K. (2023). Empowering education through generative AI: Innovative instructional strategies for tomorrow's learners. *International Journal of Business, Law, and Education*, *4*(2), 618-625. https://doi.org/10.56442/ijble.v4i2.215

Kashive, N., Powale, L., & Kashive, K. (2021). Understanding user perception toward artificial intelligence (AI) enabled e-learning. *The International Journal of Information and Learning Technology*, *38*(1), 1-19. https://doi.org/10.1108/IJILT-05-2020-0090

Latif, E., Mai, G., Nyaaba, M., Wu, X., Liu, N., Lu, G., Li, S., Liu, T., & Zhai, X. (2024). AGI: Artificial general Intelligence for education. *arXiv*, 1-19. https://doi.org/10.48550/arXiv.2304.12479

Lee, G.-G., Latif, E., Wu, X., Liu, N., & Zhai, X. (2024). Applying large language models and chain-of-thought for automatic scoring. *Computers and Education: Artificial Intelligence*, *6*, 1-15. https://doi.org/10.1016/j.caeai.2024.100213

Lin, M. P.-C., Chang, D. H., & Winne, P. H. (2025). A proposed methodology for investigating student-chatbot interaction patterns in giving peer feedback. *Educational Technology Research and Development*, *73*(1), 353-386. https://doi.org/10.1007/s11423-024-10408-3

Liu, X. (2021). Model optimization techniques for embedded artificial intelligence. *2021 2nd International Conference on Computing and Data Science (CDS)*, pp. 1-6. https://doi.org/10.1109/CDS52072.2021.00008

Mahajan, C., Krishnan, A., Nambiar, M., & Singhal, R. (2022). Hetero-rec: Optimal deployment of embeddings for high-speed recommendations. *AIMLSystems '22: Proceedings of the Second International Conference on AI-ML Systems*. *11*, 1-9. https://doi.org/10.1145/3564121.3564134

Muhammad Rahaji, J., & Sudianto, S. (2023). Implementation of chatbot for merdeka belajar kampus merdeka program using long short-term memory. *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, *12*(2), 253-262. https://doi.org/10.23887/janapati.v12i2.58794

Nasa-Ngium, P., Nuankaew, W. S., & Nuankaew, P. (2023). Analyzing and tracking student educational program interests on social media with chatbots platform and text analytics. *International Journal of Interactive Mobile Technologies (iJIM)*, *17*(05), 4-21. https://doi.org/10.3991/ijim.v17i05.31593

Nguyen, B.-S., Le, D. T., Vu, H. M., Nguyen, T. A. D., Nguyen, M.-T., & Le, H. (2022). Improving document image understanding with reinforcement finetuning. *arXiv*. http://arxiv.org/abs/2209.12561

Popescu, R.-I., Sabie, O. M., & Trușcă, M. I. (2023). The contribution of artificial intelligence to stimulating the innovation of educational services and university programs in public administration. *Transylvanian Review of Administrative Sciences*, *19*(70), 85-108. https://doi.org/10.24193/tras.70E.5

Purchase, S., Zhao, A., & Mullins, R. D. (2022). Revisiting embeddings for graph neural networks. *arXiv*. http://arxiv.org/abs/2209.09338

Ramakrishnan, R., Thangamuthu, P., Nguyen, A., & Gao, J. (2024). Revolutionizing campus communication: NLP-powered university chatbots. *International Journal of Advanced Computer Science and Applications*, *15*(6), 1-12. https://dx.doi.org/10.14569/IJACSA.2024.0150606

Ranjit, M., Ganapathy, G., Manuel, R., & Ganu, T. (2023). Retrieval augmented chest X-Ray Report generation using openAI GPT models. *arXiv*. http://arxiv.org/abs/2305.03660

Rasheed, Z., Ghwanmeh, S., & Abualkishik, A. Z. (2023). Harnessing Artificial intelligence for personalized learning: A systematic review. *Data and Metadata*, *2*(146), 1-15. https://doi.org/10.56294/dm2023146

Rios-Campos, C., Cánova, E. S. M., Zaquinaula, I. R. A., Zaquinaula, H. E. A., Vargas, D. J. C., Peña, W. S., Idrogo, C. E. T., & Arteaga, R. M. Y. (2023). Artificial intelligence and education. *South Florida Journal of Development*, *4*(2), 641-655. https://doi.org/10.46932/sfjdv4n2-001

Saetia, C., & Vateekul, P. (2018). Enhance accuracy of hierarchical text categorization based on deep learning network using embedding strategies. *2018 15th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, pp. 1-6. http://dx.doi.org/10.1109/jcsse.2018.8457376

Sajja, R., Sermet, Y., Cikmaz, M., Cwiertny, D., & Demir, I. (2024). Artificial intelligence-enabled intelligent assistant for personalized and adaptive learning in higher education. *Information*, *15*(10), 1-23. https://doi.org/10.3390/info15100596

Sajja, R., Sermet, Y., Cwiertny, D., & Demir, I. (2023). Platform-independent and curriculum-oriented intelligent assistant for higher education. *International Journal of Educational Technology in Higher Education*, *20*(42), 1-24. https://doi.org/10.1186/s41239-023-00412-7

Santos, G. A., de Andrade, G. G., Silva, G. R. S., Duarte, F. C. M., Costa, J. P. J. D., & de Sousa, R. T. (2022). A conversation-driven approach for chatbot management. *IEEE Access*, *10*, 8474-8486. https://doi.org/10.1109/ACCESS.2022.3143323

Seraquive, J. E. G., Álvarez-Muñoz, P., Palacios-Zamora, K., Gamboa, D. A. P., & Faytong-Haro, M. (2024). Unveiling the core constructs: A statistical approach to evaluating user experience with chatbots in higher education (A case study from a university in Ecuador). *Journal of Infrastructure Policy and Development 2024*, *8*(10), 1-14. https://doi.org/10.24294/jipd.v8i10.6381

Shams, G., Kim, K. K., & Kim, K. (2024). Enhancing service recovery satisfaction with chatbots: The role of humor and informal language. *International Journal of Hospitality Management*, *120*, 103782. https://doi.org/10.1016/j.ijhm.2024.103782

Siriwardhana, S., Weerasekera, R., Wen, E., Kaluarachchi, T., Rana, R., & Nanayakkara, S. (2023). Improving the domain adaptation of retrieval augmented generation (RAG) models for open domain question answering. *Transactions of the Association for Computational Linguistics*, *11*, 1-17. https://doi.org/10.1162/tacl_a_00530

Siriwardhana, S., Weerasekera, R., Wen, E., & Nanayakkara, S. (2021). Fine-tune the entire RAG architecture (including DPR retriever) for question-answering. *arXiv*. http://arxiv.org/abs/2106.11517

Sivaraman, V., Wu, Y., & Perer, A. (2022). Emblaze: Illuminating machine learning representations through interactive comparison of embedding spaces. *The 27th International Conference on Intelligent User Interfaces*, pp. 418-432. https://doi.org/10.1145/3490099.3511137

Su , J., & Yang , W. (2023). Unlocking the power of chatgpt: A framework for applying generative ai in education. *ECNU Review of Education*, *6*(3), 355-366. https://doi.org/10.1177/20965311231168423

Sugianto, S., Abidin, Z., Putra, A., & Budiman, K. (2021). Knowledge management system in a higher education institution: Development of an expertise search system. *Journal of Physics: Conference Series*, *1918*, 1-5. https://doi.org/10.1088/1742-6596/1918/4/042021

Suryanto, T. L. M., Wibawa, A. P., Hariyono, H., & Nafalski, A. (2023). Evolving conversations: A review of chatbots and implications in natural language processing for cultural heritage ecosystems. *IJRCS: International Journal of Robotics and Control Systems*, *3*(4), 955-1006. https://doi.org/10.31763/ijrcs.v3i4.1195

TonbuloĞLu, B. (2023). An evaluation of the use of artificial intelligence applications in online education. *Journal of Educational Technology and Online Learning*, *6*(4), 866-884. https://doi.org/10.31681/jetol.1335906

Tu, S., Zhang, Z., Yu, J., Li, C., Zhang, S., Yao, Z., Hou, L., & Li, J. (2023). LittleMu: Deploying an online virtual teaching assistant via heterogeneous sources integration and chain of teach prompts. *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, pp. 4843-4849. https://doi.org/10.1145/3583780.3615484

Vázquez-Cano, E., Mengual-Andrés, S., & López-Meneses, E. (2021). Chatbot to improve learning punctuation in Spanish and to enhance open and flexible learning environments. *International Journal of Educational Technology in Higher Education*, *18*(33), 1-20. https://doi.org/10.1186/s41239-021-00269-8

Walker, N. T., Ultes, S., & Lison, P. (2023). Retrieval-augmented neural response generation using logical reasoning and relevance scoring. *arXiv*. http://arxiv.org/abs/2310.13566

Wang, K., Ramos, J., & Lawrence, R. (2023). ChatEd: A chatbot leveraging chatgpt for an enhanced learning experience in higher education. *arXiv*. http://arxiv.org/abs/2401.00052

Zhang, L., & Hu, Y. (2021). A fine-tuning approach research of pre-trained model with two stage. *2021 IEEE International Conference on Power Electronics, Computer Applications (ICPECA)*, pp. 905-908. https://doi.org/10.1109/ICPECA51329.2021.9362566

Zhang, X., Sun, M., Liu, J., & Li, X. (2021). Optimal embedding calibration for symbolic music similarity. *arXiv*. http://arxiv.org/abs/2103.07656

Zhang, Z., Fang, M., & Chen, L. (2024). RetrievalQA: Assessing adaptive retrieval-augmented generation for short-form open-domain question answering. *arXiv*. http://arxiv.org/abs/2402.16457