This version of the contribution has been accepted for publication, after peer review (when applicable) but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: http://dx.doi.org/10.1007/978-3-031-93415-5_15. Use of this Accepted Version is subject to the publisher's Accepted Manuscript terms of use https://www.springernature.com/gp/open-research/policies/accepted-manuscript-terms

A virtual support agent for university students powered by a large language model: conversational user experience design considerations

Joseph Benjamin Ilagan, Wolverix Skyler Yu, Samantha Mae See, and Stephanie Rayco

Business Insights Laboratory for Development, Ateneo de Manila University, Quezon City, Philippines
jbilagan@.ateneo.edu,
{wolverix.yu,samantha.see,stephanie.rayco}@student.ateneo.edu,
WWW home page: https://www.ateneo.edu/

A virtual support agent for university students powered by a large language model: conversational user experience design considerations

Abstract. This paper presents a pilot study of a university virtual support agent system powered by a Large Language Model (LLM) engineered to address student inquiries regarding processes and policies involving a student exchange program. Traditional approaches to implementing chatbots in university student support services have had their shortcomings. This study investigates the enablement of conversational support using generative artificial intelligence and natural conversation properties inherent in LLMs and how they may overcome such shortcomings from earlier attempts at conversational support. The LLMs' susceptibility to 'hallucination' is mitigated through a combined approach of few-shot learning, retrieval augmented generation (RAG), and chain of thought (CoT) libraries in the training phase. The system utilizes OpenAI's GPT-4o model behind a Python-based web framework, enabling the conversational interface used by the students. The information system is grounded in earlier email correspondences from program heads to students, PDF files containing primers, and other sources. The pilot implementation of the system involves an outbound student exchange program in a private university in the Philippines. Sophomore students were invited to test and evaluate the system and interviewed after the chatbot sessions.

Keywords: Conversational UX · Large language models · LLM · RAG · Chatbots · University chatbots · Generative AI · GenAI.

1 Introduction

Conversational agents have been used as alternative means of providing university support services by providing a more natural and realistic user experience

(UX) [10]. Some of these conversational agents come as *chatbots* communicating automatically with a user in intelligence levels through natural language. Chatbots have been used to reduce service costs and handle several inquiries simultaneously [4]. For many administrative policies and procedures, however, the complex application process—with varying requirements and deadlines—often leads to confusion and missed deadlines. Fragmented information heightens this challenge, as the absence of a unified interface for essential details, such as administrative emails and official guidelines, results in repetitive inquiries that overwhelm faculty and staff. This paper builds on research into conversational support systems for student services using Large Language Models (LLMs) [5]. Unlike traditional chatbots, which struggle with scattered data and limited understanding, this pilot study investigates advanced LLM methodologies—like few-shot learning, retrieval-augmented generation (RAG) [7], and chain-of-thought (CoT) [11] reasoning—to enhance the student user experience. The research questions are: RQ1) What elements improve student access to administrative support? RQ2) How can a conversational agent effectively deliver administrative information? RQ3) How can this system alleviate the workload of university administrators? In addressing these questions, the study aims to contribute knowledge on LLM-driven university student support systems, focusing on user experience and administrative efficiency.

2 Review of Related Work

2.1 Chatbot Usability, Effectiveness, and User Experience in Administrative Settings

Chatbots have become essential tools in administrative settings, streamlining operations, improving response times, and enhancing user accessibility. They assist with student admissions, course registration, and financial aid processing, reducing administrative workload while providing real-time responses to user inquiries [1]. Chatbot usability studies highlight effectiveness, efficiency, and user satisfaction as key performance metrics, with improvements in natural language processing (NLP) and retrieval-augmented generation (RAG) significantly enhancing chatbot accuracy and reliability [9]. AI-powered chatbots in public administration services have also demonstrated immense potential, particularly in automating repetitive tasks, improving service accessibility, and supporting decision-making processes [8]. Unlike traditional rule-based chatbots, AI-driven models leverage deep learning and NLP to interpret complex queries, personalize responses, and continuously learn from user interactions. However, challenges such as limited training data, difficulty in understanding varied user inputs, and trust issues remain barriers to widespread adoption [8]. Addressing these limitations through enhanced data structuring, sentiment analysis, and adaptive learning models can further improve chatbot usability and effectiveness in administrative settings. Continuous refinement and integration with institutional knowledge systems will be critical for maximizing chatbot impact in both university administration and broader public service applications [1][8][9].

2.2 Generative AI, LLMs, and Conversational Interfaces

Retrieval-augmented generation (RAG) [7] is an approach that endows pre-trained generative models such as LLMs with a non-parametric memory fine-tuning approach using vector-indexed source documents. The RAG approach has been reported to have state-of-the-art performance vs. earlier parametric memory-based approaches [7].

The integration of LLMs with RAG techniques allows for more effective handling of user queries by leveraging vast datasets to provide contextually appropriate and informative responses. This capability is particularly beneficial in administrative settings, where the need for precise and timely information is paramount. Studies indicate that LLMs equipped with RAG can outperform traditional chatbots by offering enhanced user experiences through improved dialogue management and contextual understanding [3]. Moreover, the application of generative AI in conversational interfaces extends beyond simple query resolution. It facilitates personalized interactions by adapting responses based on user behavior and preferences, thus fostering a more engaging user experience. However, challenges such as data privacy concerns, the potential for biased outputs, and the necessity for continuous learning remain critical considerations in the deployment of these advanced systems.

To address these challenges, ongoing research focuses on developing robust frameworks that emphasize ethical AI usage, transparency in decision-making processes, and mechanisms for user feedback integration. These efforts aim to enhance the usability and effectiveness of generative AI applications in various domains, including education and public administration [**?**]. As generative AI continues to evolve, its role in shaping conversational interfaces will likely expand, offering new opportunities for innovation in user interaction and service delivery.

3 Architecture, Design, and Methodology

The virtual support agent was designed to emulate the experience of talking to the human coordinator for outbound student exchange about procedures and guidelines for applying to the program. The virtual support agent thus required the following components loosely based on [2], [4], and [6]:

- An ability to converse with humans naturally and in a human-like manner.
- A chat-based user interface.
- Access to up-to-date and accurate information on university policy.

The chatbot's architecture is a distributed system that seamlessly integrates a Django-based web application and a Telegram bot to provide natural language processing capabilities powered by OpenAI's GPT-4o model. This architecture is designed to balance computational efficiency, retrieval accuracy, and response relevance, making it an effective tool for managing user queries and handling diverse datasets.

The Django application serves as the back-end for file management, embedding generation, and document storage. When users upload files, the system processes these files to generate embeddings using OpenAI's text embedding model. To enhance the retrieval-augmented generation (RAG) process, the system employs a chunking strategy. This strategy divides large documents into smaller, manageable segments based on predefined markers, such as paragraphs or headings. The chunking ensures that each segment contains coherent and contextually complete information while staying within the token limit constraints of the embedding model. The resulting embeddings, along with metadata (e.g., document names or chunk indices), are stored in a CSV file for later retrieval.

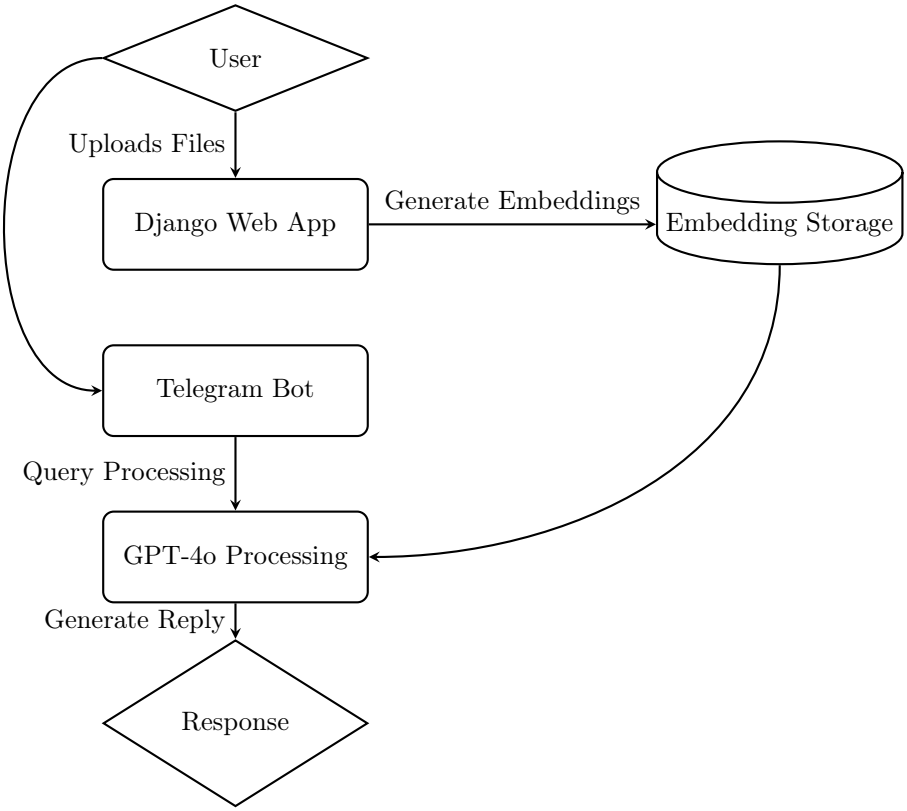


Diagram 1: "System Architecture Overview"

The Telegram bot, developed using the aiogram library, serves as the primary interface for user interactions. It leverages OpenAI's GPT-4o model for natural language processing, enabling it to understand and generate responses to user queries. The bot follows a retrieval strategy that integrates text embeddings into its workflow. Upon receiving a query, the bot retrieves the most relevant chunks of information from the preprocessed documents. This ensures that the model

generates responses based on accurate and contextually appropriate information. The bot's design incorporates several advanced natural language processing techniques:

1. Few-shot learning: The bot includes example Q&A pairs in its system prompts to guide GPT-4o in understanding the task context and generating coherent, high-quality responses. These examples provide additional context, ensuring more accurate query handling.

2. Chain of Thought (COT) reasoning: The bot employs a step-by-step reasoning framework, especially for complex queries. By breaking down such queries into intermediate steps, the bot provides more structured and transparent responses. This approach enhances the bot's ability to tackle intricate questions while maintaining clarity and precision.

3. Optimization for token usage: The system significantly reduces token consumption without compromising response quality by focusing on retrieving only the most relevant information. This optimization minimizes costs and ensures faster response times.

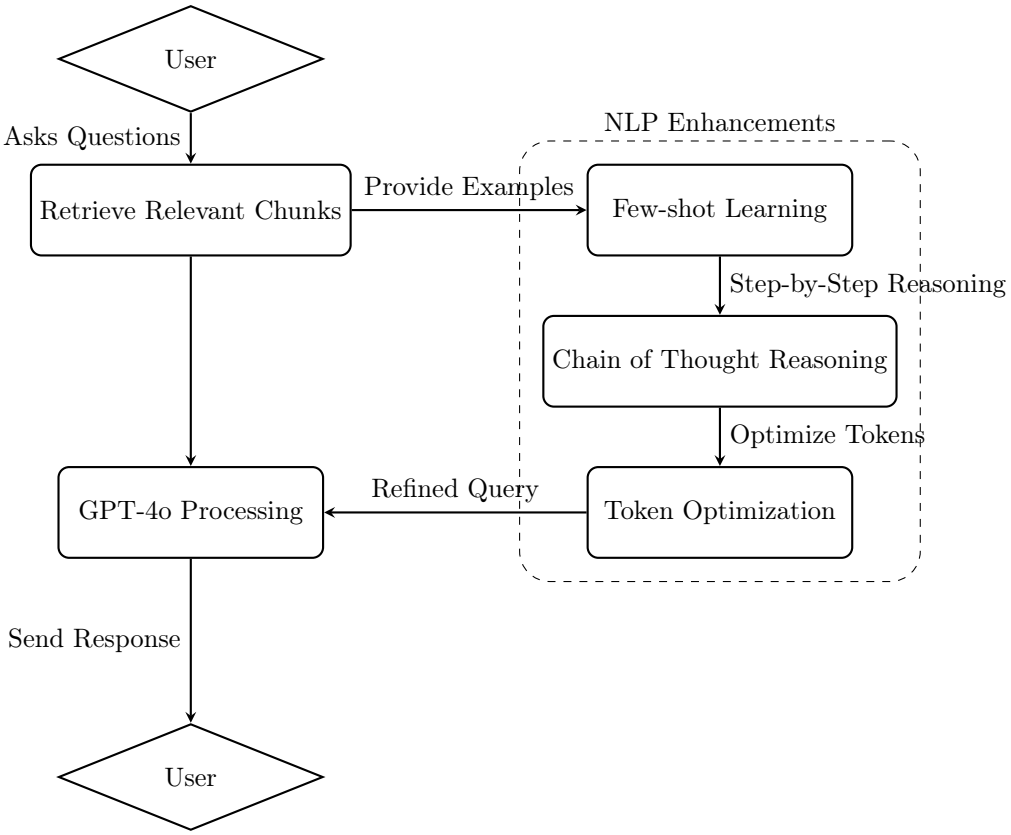


Diagram 2: "Query Processing Workflow"

To mitigate the risks of hallucinations (where the model generates plausible-sounding but inaccurate information), the bot relies on robust retrieval mechanisms and structured prompts. Embedding-based retrieval ensures that the chatbot accesses factual, document-based information, while the system prompts clearly outline the task boundaries and expected outputs. The interplay between retrieval-augmented generation and system-level constraints ensures that the chatbot delivers context-aware, relevant, and accurate responses.

This distributed architecture supports scalability and modularity. The decoupling of the backend file management system and the user-facing bot allows independent updates or improvements to either component without disrupting the entire workflow. Future enhancements, such as integrating additional NLP models or expanding storage capabilities, can be implemented with minimal architectural changes.

The chatbot is fed with various pieces of information to enhance its knowledge and response capabilities. Primarily, it uses embedded text from uploaded files, which are processed and stored in a CSV file. These embeddings contain information about the international outbound student exchange program program called "Junior Term Abroad" (JTA) at a private university in the Philippines, derived from email correspondences, PDF files, and other sources. Application guidelines, requirements, and FAQ's were some of the files uploaded. User information is also maintained to personalize responses, including name, school, and course, which are collected during the initial bot setup. Conversation history is tracked and incorporated into queries (), allowing for context-aware responses. The bot is also guided by system prompts that define its behavior and knowledge domain. These prompts include few-shot learning examples to help the bot understand the types of questions it might encounter and how to respond appropriately.

Participants were recruited via a Google Forms survey, disseminated by batch representatives, faculty members, and student organization leaders to target third-year students from various academic schools. Respondents were then stratified into four confidence tiers, ranging from those needing comprehensive assistance to those fully confident with the provided materials. This stratification allowed for a more nuanced assessment of student readiness and helped identify specific areas where additional support was required.

The evaluation process that followed after was a two-phase approach. The first phase involved a pre-test survey designed to capture both quantitative and qualitative insights into students' familiarity with the application process, their commitment to the JTA program, and any knowledge gaps. These insights were instrumental in refining the chatbot's functionality, ensuring it addressed students' challenges effectively. In the second phase, post-test interviews gathered qualitative feedback on the chatbot's usability, intuitiveness, and response accuracy. Participants also compared the chatbot to traditional support methods, such as consulting faculty or using online resources, providing a broader eval-

uation of its effectiveness. Finally, suggestions for improvement were gathered to inform future iterations of the chatbot, ensuring it becomes an even more reliable tool for guiding student applications.

4 Results

The pilot study yielded valuable insights into the chatbot's effectiveness, highlighting both its strengths and areas for improvement. Participants generally praised the chatbot's intuitive interface and its ability to provide immediate, fact-based responses to frequently asked questions. The anonymity offered by the chatbot was particularly appreciated by introverted students, who found it to be a non-judgmental and accessible alternative to traditional consultation methods. The chatbot was especially effective in addressing routine inquiries related to deadlines, application requirements, and financial considerations, which were previously covered in information sessions or documentation but often overlooked by students.

Despite these positive outcomes, the chatbot exhibited notable limitations in handling complex, subjective queries. Specifically, it struggled with questions related to personalized experiences, such as exchange program budgeting, which varies based on individual spending habits, and resume guidance, which requires nuanced, context-specific advice. The chatbot's inability to provide personalized responses in these areas resulted in user frustration, as its fallback message—directing students to the exchange coordinator—was perceived as an unsatisfactory solution. However, to mitigate this issue, a backend system was developed, allowing administrators to monitor recurring queries and update the chatbot's database accordingly.

Another area for improvement identified was the chatbot's response to queries that required dynamic data interpretation. While the chatbot could provide static, pre-determined answers efficiently, it faced challenges when responding to inquiries that involved subjective decision-making or complex administrative policies, such as course crediting across different partner universities. Participants also suggested that features such as personalized checklists, interactive data visualization, and deeper integration with financial planning tools could enhance the chatbot's utility.

Overall, the pilot study demonstrated that the chatbot significantly reduced administrative workload by handling repetitive inquiries and offering immediate access to relevant information. Participants acknowledged the chatbot's potential in improving the efficiency of the exchange program process but emphasized the need for expanded data sources and improved personalization features to enhance user experience and engagement.

5 Discussion and Conclusion

The implementation of the chatbot in the student exchange program has proven to be an effective tool for enhancing administrative efficiency and student engage-

ment. By automating responses to routine inquiries, the chatbot has significantly reduced the workload of administrative staff, allowing them to focus on strategic tasks such as program development and student support initiatives. The chatbot's ability to centralize essential resources—such as course equivalency lists, financial breakdowns, and application deadlines—streamlines the dissemination of critical information, reducing the need for manual distribution and repetitive clarifications.

A key advantage of the chatbot is its accessibility and immediate responsiveness, which contrasts with the delays associated with traditional support channels such as faculty consultations and email exchanges. Students reported that responses via traditional methods often took several days, with some inquiries going unanswered if staff deemed the information readily available in existing documents. The chatbot effectively addresses this gap by offering instant access to commonly sought-after details, improving response accuracy and user satisfaction. Furthermore, its 24/7 availability offers a significant advantage, especially for prospective students and their parents who require immediate information for decision-making purposes.

However, the study also highlighted areas where the chatbot's capabilities need enhancement. Its struggle with subjective and complex inquiries—such as budgeting advice and unique exchange experiences—suggests the need for improved data enrichment strategies. In response to this challenge, redirecting students to alumni events and testimonials has been identified as a viable solution, providing first-hand experiences that the chatbot currently cannot replicate. Additionally, the chatbot's limited capacity to interpret subjective data, such as individualized course credit assessments, underscores the importance of integrating advanced AI features, such as natural language processing (NLP) models capable of contextual understanding and personalization.

Scalability remains a crucial consideration for the future of the chatbot. Encouraged by the initial success of the pilot, the university is exploring opportunities to expand its application to other departments facing a high volume of inquiries. Two additional departmental leads have already expressed interest in adopting the technology, indicating the broader potential for AI-driven administrative support. Future improvements will focus on integrating the chatbot with additional data sources, such as CSV file processing capabilities, to facilitate better data management and comprehension.

The effectiveness of retrieval-augmented generation (RAG) and chain-of-thought (CoT) prompting techniques in mitigating hallucinations was a notable success. Despite attempts by students to elicit irrelevant or fabricated responses, the chatbot consistently maintained accuracy by relying strictly on the available knowledge base. This finding underscores the importance of grounding AI responses in verified data sources to maintain credibility and trust among users.

In conclusion, the chatbot has demonstrated significant potential in streamlining administrative processes and enhancing student autonomy in navigating the exchange application process. While it excels at providing quick, factual information, further enhancements—such as expanding its knowledge base, im-

proving personalization capabilities, and incorporating advanced data processing—will be crucial in maximizing its impact. The study's findings suggest that with continued development and institutional support, AI-powered chatbots can play a transformative role in improving efficiency, reducing administrative burden, and ultimately enhancing the overall student experience.

However, challenges such as institutional approval processes and resistance to technological change within academic environments remain potential barriers to widespread adoption. Overcoming these obstacles will require a proactive approach, including stakeholder engagement, user education, and clear demonstrations of the chatbot's tangible benefits to the institution.

Acknowledgments. We would like to thank Ateneo de Manila University through the John Gokongwei School of Management (JGSOM) and the Business Insights Laboratory for Development (BUILD) for enabling our participation and future presentation of this study.

References

1. Al-Jaf, K., Oz, C., Mahmud, H., Rashid, T.A.: Leveraging Chatbots for Effective Educational Administration: A Systematic Review (Oct 2024), https://doi.org/10.20944/preprints202410.0238.v1
2. Bartneck, C., Kulić, D., Croft, E., Zoghbi, S.: Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. International Journal of Social Robotics **1**(1), 71–81 (Jan 2009). https://doi.org/10.1007/s12369-008-0001-3, https://doi.org/10.1007/s12369-008-0001-3, 1791 citations (Crossref) [2024-10-18]
3. Bennion, M.R., Hardy, G.E., Moore, R.K., Kellett, S., Millings, A.: Usability, Acceptability, and Effectiveness of Web-Based Conversational Agents to Facilitate Problem Solving in Older Adults: Controlled Study. doi: 10.2196/16794 (May 2020)
4. Dibitonto, M., Leszczynska, K., Tazzi, F., Medaglia, C.M.: Chatbot in a campus environment: design of LiSA, a virtual assistant to help students in their university life. In: Human-Computer Interaction. Interaction Technologies: 20th International Conference, HCI International 2018, Las Vegas, NV, USA, July 15–20, 2018, Proceedings, Part III 20. pp. 103–116. Springer (2018)
5. Ilagan, J.B., Ilagan, J.R.: A prototype of a conversational virtual university support agent powered by a large language model that addresses inquiries about policies in the student handbook. Procedia Computer Science **239**, 1124–1131 (2024), https://www.sciencedirect.com/science/article/pii/S1877050924015217, publisher: Elsevier
6. Kuligowska, K.: Commercial Chatbot: Performance Evaluation, Usability Metrics and Quality Standards of Embodied Conversational Agents (Feb 2015), https://papers.ssrn.com/abstract=2569637
7. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.t., Rocktäschel, T.: Retrieval-augmented generation for knowledge-intensive nlp tasks. Advances in Neural Information Processing Systems **33**, 9459–9474 (2020), https://proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html

8. Nirala, K.K., Singh, N.K., Purani, V.S.: A survey on providing customer and public administration based services using AI: chatbot. Multimedia Tools and Applications **81**(16), 22215–22246 (Jul 2022). https://doi.org/10.1007/s11042-021-11458-y, https://doi.org/10.1007/s11042-021-11458-y
9. Ren, R., Zapata, M., Castro, J.W., Dieste, O., Acuña, S.T.: Experimentation for Chatbot Usability Evaluation: A Secondary Study. IEEE Access **10**, 12430–12464 (2022). https://doi.org/10.1109/ACCESS.2022.3145323, https://ieeexplore.ieee.org/document/9687577/?arnumber=9687577
10. Tubin, C., Mazuco Rodriguez, J.P., de Marchi, A.C.B.: User experience with conversational agent: A systematic review of assessment methods. Behaviour & Information Technology **41**(16), 3519–3529 (2022)
11. Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., Zhou, D.: Chain-of-Thought Prompting Elicits Reasoning in Large Language Models (Jan 2023), http://arxiv.org/abs/2201.11903