JayBot – Aiding University Students and Admission with an LLM-based Chatbot

Julius Odede
Ingo Frommholz
J.A.Odede3@wlv.ac.uk
ifrommholz@acm.org
University of Wolverhampton
Wolverhampton, UK

ABSTRACT

This demo paper presents JayBot, an LLM-based chatbot system aimed at enhancing the user experience of prospective and current students, faculty, and staff at a UK university. The objective of JayBot is to provide information to users on general enquiries regarding course modules, duration, fees, entry requirements, lecturers, internship, career paths, course employability and other related aspects. Leveraging the use cases of generative artificial intelligence (AI), the chatbot application was built using OpenAI's advanced large language model (GPT-3.5 turbo); to tackle issues such as hallucination as well as focus and timeliness of results, an embedding transformer model has been combined with a vector database and vector search. Prompt engineering techniques were employed to enhance the chatbot's response abilities. Preliminary user studies indicate JayBot's effectiveness and efficiency. The demo will showcase JayBot in a university admission use case and discuss further application scenarios.

CCS CONCEPTS

• **Information systems** → **Users and interactive retrieval**; • **Computing methodologies** → *Discourse, dialogue and pragmatics.*

KEYWORDS

Chatbot, Large Language Models, Machine Learning, Interactive Information Retrieval, Vector Database, Artificial Intelligence, Retrieval Augmented Generation

ACM Reference Format:
Julius Odede and Ingo Frommholz. 2024. JayBot – Aiding University Students and Admission with an LLM-based Chatbot. In *Proceedings of ACM SIGIR Conference on Human Information Interaction And Retrieval (CHIIR '24).* ACM, New York, NY, USA, 5 pages. https://doi.org/XXXXXXX.XXXXXXX

CHIIR '24, March 10-14, 2024, Sheffield, UK

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Prospective students often face difficulties in accessing quick information about courses, admissions, scholarships, and tuition. Prolonged admission wait times compel both local and international applicants to turn to university websites, where they frequently rely on live chat platforms for course details, applications, housing, and admission status. Delayed responses can frustrate students. A surge in applicant numbers can exacerbate waiting times, leading to negative reviews and a decline in student interest. Universities risk revenue losses due to waning student enthusiasm. This paper aims to rectify these issues by presenting a chatbot called *JayBot*, that can be employed for example on the websites of UK or international universities to provide swift information and enhance the prospective student experience.

Generative AI as a subset of artificial intelligence has made significant impacts on the world today and has gained immense popularity. Its ability to generate text, audio, video, images, and code using Large Language Models (LLMs) has resulted in the development of modern chatbot applications such as OpenAI's ChatGPT, Microsoft Bing, and Google Bard. These powerful chatbot applications have helped to improve businesses by acting as virtual assistants that can perform various tasks such as crunching numbers, retrieving insights from documents, managing customer queries, reducing costs, increasing revenues, and improving experiences. JayBot aims to leverage the use cases of generative AI and prompt engineering, an emerging force in AI, to build a chatbot for UK universities that will aid in solving customer support issues and providing instant responses to enquiries. The demonstrated prototype attempts to fix the hallucination and memory concise problem of LLMs which users have noticed while interacting with ChatGPT, by introducing a vector database to the model. The prototype uses prompt engineering to steer the model to provide appropriate responses to user queries (complicated or domain-specific inquiries).

2 RELATED WORK

Over the years, conversational information retrieval systems that allow users to use natural language in a dialogue to express their information needs, have gained prominence, with chatbots being one of the most known examples of conversational interfaces [11]. In [8], the authors discuss the primary role of chatbots in facilitating Human-Computer Interaction (HCI), enabling users to express their interests and questions naturally. They emphasize that chatbots simulate human conversation by combining language models and computational algorithms. Additionally, the work in [1] underscores that chatbots aim to simplify data retrieval for users,

allowing them to ask questions in natural language. They developed a conventional chatbot using Java programming. Thakkar et al. [10] used a screening method to analyse chatbots supporting learning via Facebook Messenger. Their findings were published in a web directory, demonstrating that chatbots provided effective educational support. Liebrecht and Hooijdonk [6] conducted research on educational chatbots following the PRISMA approach, adding to the knowledge on chatbot use in education. The University of California in 2021 showcased the effectiveness of an AI chatbot in addressing library inquiries, assisting with course enrolment, and guiding students to admissions offices. Chatbots offer 24/7 availability, benefiting international students regardless of time zones [2]. Recent developments in generative AI and LLMs are promising for the next generation of intelligent agents and chatbots. However, they suffer from several problems, such as making up plausible-looking yet incorrect or untimely information (commonly referred to as "hallucination"). It has been shown that augmenting answers with retrieved documents (retrieval-augmented LLMs) can reduce hallucination [9]. The prototype demonstrated here is based on this assumption.

3 TARGET USERS AND INITIAL EVALUATION

3.1 Target Users

JayBot is an AI chatbot system that provides quick responses to enquiries about a person, product, service, business, or company. The prime users of the current JayBot implementation are current or prospective students who seek information about courses and programmes at a UK university. JayBot has also been successfully deployed to the University of Wolverhampton live chat team for testing purposes, who act as a mediator to provide students with the required information.

3.2 Out-of-Context Responses and Hallucination

In a first initial evaluation, we tested ChatGPT regarding its capability to satisfy the information needs of our target users. Large Language Models (LLMs) are known for providing wrong or out-of-context information to users. For instance, when we asked the chat model (GPT-3.5) about the tuition fee of the MSc Data Science course at the University of Wolverhampton, we can see in Figure 1 that the model was unable to present a correct answer due to a lack of up-to-date and focused information. This is because it lacks access to the proprietary and niche data of that particular domain. While it can provide answers to a wide range of topics, it is not trained to handle these specific queries. In this case, it lacks access to the data of the MSc Data Science course under consideration, so it tends to hallucinate answers for topics it has no knowledge of or context for. Another reason for the model's hallucination and inability to provide a proper response is due to the inefficient knowledge updates of large language models. Generative AI models require tens of thousands of high-cost GPU training hours to retrain on up-to-date information and once the training process completes, AI models are stuck in a "frozen" version of the world they saw during training. OpenAI models such as GPT-3.5 used for ChatGPT

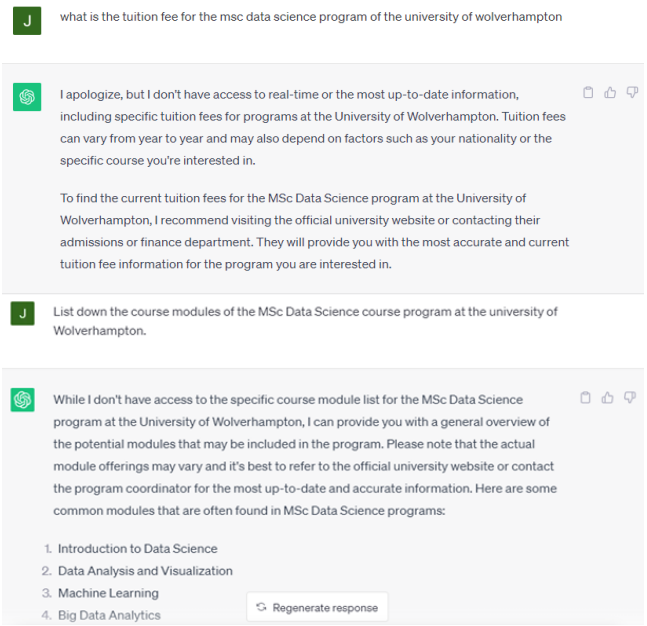

Figure 1: ChatGPT responses.

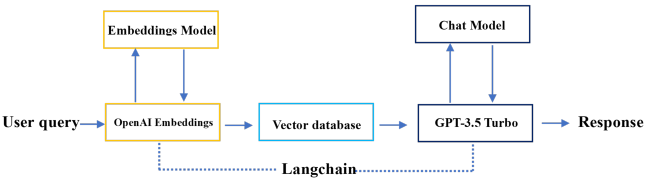

Figure 2: JayBot backend system architecture.

and GPT-3.5 turbo used for this study were trained up until September 2021 therefore these models are unable to answer queries on topics, information, and happenings after September 2021.

Our solution to these problems is to utilise a non-parametric vector database that uses word embeddings generated by the LLM to provide accurate domain-specific information.

4 JAYBOT DESIGN AND IMPLEMENTATION

4.1 Backend and Frontend

Figure 2 shows the architecture of the backend part of our system. Two OpenAI language models (chat model and embedding model) were used to build the chatbot backend. The system utilises the OpenAI GPT-3.5 turbo model. The OpenAI embedding model creates vector representations of text, to convert the user's textual input into vector embeddings for GPT-3.5 turbo to perform chat completion to users. A vector database is a type of database that stores and enables efficient retrieval of vector representations of data points. Vector databases are often used to store word embeddings, document embeddings, or other vector representations of

textual data. A vector database will be required to store the embedded version of the chatbot knowledge base (scrapped data of the MSc courses) in a vector space after conversion by the embedding model. Embeddings are generated by AI models and have many features which make their representation challenging to manage. In the context of artificial intelligence and machine learning, these features represent different dimensions of data that are essential for understanding patterns, relationships and underlying structures. That is why a vector database is chosen specifically for handling this type of data. The JayBot backend was written in Python. Flask was used to handle HTTP requests, routing, and to provide a structure for organising the backend codes. Langchain [7] was used for chaining the different components of the chatbot backend. Pinecone[1] was chosen as vector database.

The frontend of the chatbot system was built using JavaScript, Typescript, CSS, SCSS (Sassy Cascading Style Sheets). The chatbot user interface was designed using a combination of physical design elements that presented the chatbot application to be visually appealing and user-friendly. These design elements are: colour (to match the university website's colour scheme); images and graphics (to mirror the visual aesthetics of the university website); typography and visual effects (to ensure positioning and legibility of texts within the web and mobile app); icons and buttons (to enable users to perform specific actions). Fig. 4 shows the simple chat interface.

4.2 Tackling Hallucination

To tackle the hallucination problem and provide the required accurate domain knowledge, relevant knowledge was scraped from the university's website containing course information. The aforementioned embedding model was applied to create vector embeddings from the content, which are stored in the vector database with some reference to the original content the embedding was created from. The same embedding model is used when JayBot issues a query, to create query embeddings to retrieve similar vector embeddings from the database using vector search (Fig. 3). This was done so that the chatbot applications could retrieve contextually relevant and up-to-date embeddings from the vector database instead of from the model itself. The database becomes a long-term and up-to-date retrieval memory for the chatbot.

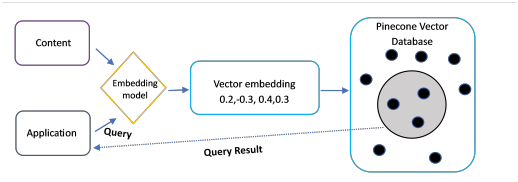

Figure 3: Vector embeddings and database responses.

4.3 Prompt Engineering

After the incorporation of the Pinecone vector database with the chatbot model, the chatbot is able to respond to users' requests through Retrieval Augmented Generation (RAG) [3] and vector search. However, instead of having a search-based and redundant system that only searches and retrieves information from a database, our goal is to have a chatbot that can engage in conversations, remember conversations, and keep the dialogue flowing with users to create a human-like conversation and experience for users. To achieve this, prompt engineering is employed.

Prompt engineering is an art or technique of telling LLMs what to do by using prompts and instructions. Interestingly, LLMs have been trained on a vast array of data on almost everything around us. With prompt engineering, we can simply instruct a model on what to do. For JayBot, we instructed the model on how to respond to users using prompts. The below prompt was used for the chatbot model at the backend to improve its response to users.

You are given a context delimited by "' along with a question. Your objective is to generate an appropriate answer. include the actual website links that contain https and contact details in your answers when necessary. Only include website links found in the database. Do not provide wrong information or wrong website links to users. Format the answer appropriately based on the user's query. Remember the context of the conversation and keep the dialogue flowing smoothly.

With a vector database now included and prompt engineering in place, if we ask the chatbot to provide information about the tuition fee of the MSc Data Science course of the same university, it fetches this information from the knowledge base of the vector database rather than from what the original GPT-3.5 turbo model saw during training. This results in the more accurate response shown in Figure 4. This example also shows how our approach restricts the chatbot to only reply to queries concerning the university's courses. When asked about a course provided by a different university, JayBot politely refuses to answer and leads the discussion back to the original university – the aim of the chatbot is not to promote a competitor's provision in a competitive higher education market as we find it in the UK and other countries.

5 EVALUATION AND IMPACT

A preliminary evaluation of JayBot's performance was conducted. During the testing stage, participants were given a questionnaire and asked to evaluate the chatbot application. The underlying use case was the search for information about a specific university course. Participants were asked to use the university's live chat, where they can chat with a staff member or student, and to provide feedback on their experience in a short survey. Later, they were asked to use JayBot for the same information needs. Participants were given a questionnaire and asked to evaluate the chatbot application. The results of the evaluation are summarised in Table 1[2]. Unsurprisingly, the chatbot replied promptly and was always available, compared to a human in the live chat, who might not be immediately available. But it also seems JayBot is ahead of a human-conducted live chat when it comes to ease of use, quality of support and overall user satisfaction, which is not necessarily an expected, yet promising, result.

During our trial runs, JayBot has helped the admission team at the University of Wolverhampton to reduce the workload of chat

[1] https://www.pinecone.io/

[2] More detailed results can be found at https://github.com/Greenconsult/Jaybot_Chiir_Paper_Evaluation and https://github.com/Greenconsult/Jaybot-Chiir-Paper, respectively.

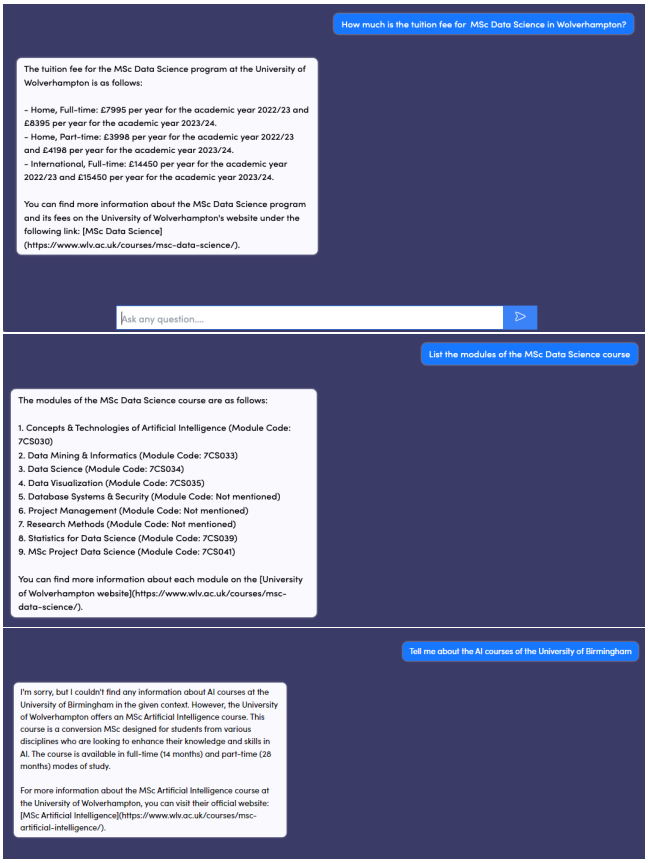

Figure 4: JayBot's more accurate and focused responses.

Metric	Live Chat (40 responses)	JayBot (50 responses)
Average response time	8.5 hours	5 seconds
Availability of support	62% complained about no human presence on the live chat.	Almost all users received responses on the chatbot app within 5 seconds.
Quality of support	28% received the information they were looking for on the live chat platform, but possibly delayed or insufficiently.	More than 90% of users find the information from the chatbot to be completely accurate.
Ease of use	50% find the registration process before chatting on the live chat to be time-consuming, while around a quarter of users find it to be stressful. Only around a quarter find the process to be simple.	Around 85% of users find the chatbot application to be very easy to use. Users get on board without any registration hassle.
User satisfaction	Around three quarters (72%) of users rated their experience on the live chat bad or very bad. Only 28% of users had a good experience.	96% of users were very satisfied with their experience on the chatbot application.

Table 1: JayBot compared to traditional live chat (summary).

messages and call inquiries. This indicates that JayBot can have a huge impact during peak admission times and clearing periods. We expect that providing prospective students with relevant information in a timely manner even during busy periods will have a positive impact on admissions and also on student satisfaction. JayBot can be extended to function as a conversational assistant that aids students throughout their life on campus. Furthermore, JayBot can be used by all industries to provide quick, fast information about businesses and products. For example, in commercial industries, it can be used for selling products, product recommendations, and providing fast information about the product and services (with the potential to scale the growth of small businesses). In health industries, JayBot can tell a patient the nearest healthcare facility to visit and also the contact details of the health facility. It can be used in enterprise search to help find relevant information in heterogeneous data repositories within the intranet [5].

6 DEMONSTRATION

We will demonstrate JayBot by showcasing a university admissions use case where prospective students are looking for information about a specific course. This includes the content of the course, registration fees, internship opportunities but also general questions such as about available stipends or accommodation[3].

7 CONCLUSION AND FUTURE WORK

In this paper we presented JayBot, which utilises GPT-3.5 embeddings to create a chatbot for prospective and current university students, faculty staff and admission, to satisfy their information needs about courses they are interested in and enhance their user experience while searching. The problem of out-of-context knowledge and hallucination was tackled by utilising a vector database containing embeddings based on the user input and material about the respective courses. The response and dialogue flow of the chatbot was improved using prompt engineering. Based on the findings of this paper, chatbots have the potential to be a cost-effective solution that can benefit multiple parties within a university community. By improving user experience and reducing the workload of students and personnel, generative AI applications such as chatbots can significantly impact the operations of a university. It can be concluded that their use holds great promise for enhancing efficiency and reducing operational burden.

Future work can address several facets of the system and its application. Further extensive user studies should be conducted to provide some more robust evaluation. Different LLMs such as GPT-4, BARD or existing open source models might be integrated and evaluated. The chatbot can be upgraded to a learning and research tool, by adding the full content of the modules of university programs into the chatbot knowledge base for learning purposes, and by incorporating an academic search engine for researching articles. The chatbot responses can also be upgraded to include images and emojis in its responses to users. The admission portal of the universities in the UK could be integrated into the backend of the chatbot system so that users can get real-time feedback on their admission status by chatting with the bot. JayBot is not bound to UK universities or higher education institutions. When applied as a chatbot and retrieval system to other venues in different countries, cultural aspects might be considered in the chatbot design [4].

ACKNOWLEDGMENTS

This work was funded by the European Union under the Horizon Europe grant OMINO – Overcoming Multilevel INformation Overload (grant number 101086321, http://ominoproject.eu), guarantee funded by UKRI (ref EP/X040496/1).

[3] A short video demonstrating JayBot can be accessed at https://drive.google.com/file/d/1W7AtHj5N2YD4TJ4r1by6kAN4xV714-CE/view?usp=sharing and JayBot is available at https://chat-bot-ui-lovat.vercel.app.

REFERENCES

[1] A. Androutsopoulou, N. Karacapilidis, E. Loukis, and Y. Charalabidis. 2019. Transforming the communication between citizens and government through AI-guided chatbots. *Government Information Quarterly* 36, 2 (2019), 358–367. https://doi.org/10.1016/j.giq.2018.10.001

[2] I. Bunosso and L. R. Levine. 2023. The Influence of Chatbot Humor on Consumer Evaluations of Services. *International Journal of Consumer Studies* (2023), 545–562. https://doi.org/10.1111/ijcs.12849

[3] Deng Cai, Yan Wang, Lemao Liu, and Shuming Shi. 2022. Recent Advances in Retrieval-Augmented Text Generation. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Madrid Spain, 2022-07-06). ACM, 3417–3419. https://doi.org/10.1145/3477495.3532682

[4] Karen Chessum, Haiming Liu, and Ingo Frommholz. 2022. A Study of Search User Interface Design Based on Hofstede's Six Cultural Dimensions. In *Proceedings 6th International Conference on Computer-Human Interaction Research and Applications (CHIRA 2022)* (Valetta, Malta). SciTePress, 145–154. https://doi.org/10.5220/0011528700003323

[5] Udo Kruschwitz and Charlie Hull. 2017. Searching the Enterprise. 11, 1 (2017), 1–142. https://doi.org/10.1561/1500000053

[6] C. Liebrecht and C. V. Hooijdonk. 2020. *Creating Humanlike Chatbots: What Chatbot Developers Could Learn from Webcare Employees in Adopting a Conversational Human Voice.* Springer International Publishing. https://doi.org/10.1007/978-3-030-39540-7

[7] Richard MacManus. 2023. *LangChain: The Trendiest Web Framework of 2023, Thanks to AI.* The New Stack. https://thenewstack.io/langchain-the-trendiest-web-framework-of-2023-thanks-to-ai/ Last visited 31/10/2023.

[8] L. Nurul, H. Mudofi, and W. Yuspin. 2023. Evaluating Quality of Chatbots and Intelligent Conversational Agents of BCA' (Vira) Line. *Vira Line* (2023), 532–542.

[9] Kurt Shuster, Spencer Poff, Moya Chen, Douwe Kiela, and Jason Weston. 2021. Retrieval Augmentation Reduces Hallucination in Conversation. In *Findings of the Association for Computational Linguistics: EMNLP 2021* (Punta Cana, Dominican Republic). Association for Computational Linguistics, 3784–3803. https://doi.org/10.18653/v1/2021.findings-emnlp.320

[10] J. Thakkar, P. Raut, Y. Doshi, and K. Parekh. 2018. Erasmus-AI Chatbot. *International Journal of Computer Sciences and Engineering* 6 (2018). https://doi.org/10.26438/ijcse/v6i10.498502

[11] Hamed Zamani, Johanne R Trippas, Jeff Dalton, and Filip Radlinski. 2023. Conversational information seeking. *Foundations and Trends® in Information Retrieval* 17, 3-4 (2023), 244–456.