Proceedings of the 4th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/46/20241435

# Chatbot for retrieving information about Chinese universities

**Tiankai Wang**

Department of Vehicle Engineering, Chang'an University, Xi'an, China

2022900349@chd.edu.cn

**Abstract.** In this research endeavor, the creation of a highly capable and adaptable educational tool has been undertaken, aimed at simplifying the complex process of university selection. A primary achievement achieved thus far is the development of a chatbot that excels in delivering precise and informative responses to user queries concerning Chinese universities. However, the vision extends beyond this initial accomplishment. The aspiration is to empower this chatbot with the ability to recommend universities tailored to users' specific needs, incorporating factors such as academic performance, hobbies, and areas of expertise. This holistic approach to educational guidance aims to revolutionize the way students make one of the most crucial decisions in their academic journeys. In conclusion, this project marks the genesis of a powerful and versatile university selection tool with the potential to significantly alleviate the burden on students and parents when it comes to choosing the right educational institution. Future work will focus on implementing advanced recommendation algorithms, leveraging machine learning, and data analytics to create an intelligent and highly personalized advisor. The ultimate goal is to make this chatbot an indispensable asset for all students embarking on their educational pursuits, providing tailored recommendations for a brighter academic future.

**Keywords:** Chatbot, NPL, University.

## 1. Introduction

In the digital age, access to information has become increasingly crucial, especially for individuals seeking higher education opportunities [1-3]. This is particularly significant in the context of China, where a unique and highly competitive education system exists. Annually, approximately ten million high school seniors participate in the National College Entrance Examination (NCEE). These students invest a significant portion of their time and energy in preparing for the NCEE, focusing on mastering the high school curriculum. However, this rigorous preparation leaves them with limited time and energy to explore and understand their future university options before the NCEE. In the immediate aftermath of the examination, which takes less than a month to release results, most students choose to engage in activities they enjoy alleviating the pressure they endured during their high school years. It is only in the week following the release of their scores that they must make crucial decisions regarding their university choices. During this short timeframe, students are faced with the daunting task of selecting a university without in-depth knowledge about the options available. Many find themselves lacking comprehensive information about the universities they are considering, and this lack of knowledge extends even after they have entered their chosen institutions. This gap in understanding and information accessibility presents a significant challenge for students in making informed decisions about their higher education journey.

DOI: 10.54254/2755-2721/46/20241435

Moreover, the impact of this limited information access goes beyond just the initial university selection process. It often continues throughout the students' academic journeys, affecting their overall satisfaction and engagement with their chosen universities. Students who are not well-informed about their institutions may miss out on valuable opportunities, such as participation in research projects, specialized courses, or extracurricular activities that align with their interests and career aspirations. Recognizing the critical importance of addressing this information gap, the research seeks to provide a holistic solution that not only assists students in making informed initial choices but also supports them throughout their university experiences. By leveraging the capabilities of the intelligent chatbot, this study aims to empower students with the knowledge they need to thrive academically and personally during their higher education journey.

The existing landscape of university information resources primarily consists of static websites, brochures, and official documents, which often lack the user-friendliness and interactivity that modern students expect. This research aims to fill this research gap by introducing a dynamic and conversational chatbot that not only offers information but engages users in meaningful conversations to understand their unique preferences and requirements.

This research addresses information accessibility and user engagement in university information retrieval. Traditional methods have limitations for fast-paced decision-making by Chinese high school graduates. The chatbot offers [4-6] personalized recommendations and in-depth insights into universities to bridge this gap. This study aims to create an engaging, user-centric experience guiding students in their university selection journey.

In subsequent sections, this study detailed the methodologies, present chatbot performance results, discuss implications, and conclude with insights for advancing chatbot technology in Chinese university information retrieval. This empowers students to make informed higher education choices.

## 2. Method

### *2.1. Dataset preparation*

The foundation of this research lies in the dataset used to train and evaluate the chatbot for retrieving information about Chinese universities. The dataset is sourced from two primary and reputable sources including the Shanghai Ranking Academic Ranking of World Universities and the Time Guanghua's Chinese university ranking dataset in CSV format [7]. These sources are renowned for providing comprehensive information about universities in China.

Each data point in the dataset includes extensive background information about a specific university. This information encompasses details such as the university's name, location, academic ranking, areas of expertise, and historical context. The dataset also includes data on the university's faculty, student population, and notable achievements.

To prepare the dataset for training, comprehensive preprocessing steps were carried out. These steps encompassed the import and structuring of the data into a format suitable for training the chatbot model. The resulting dataset is then stored in a vector store using FAISS, ready for further processing and model training. This dataset preparation procedure ensures that the chatbot has access to well-structured and informative data, enabling it to furnish accurate and pertinent information regarding Chinese universities. In the subsequent sections of this paper, the chatbot model's architecture, its functionalities, and the results and discussions concerning its performance in aiding users in retrieving information about Chinese universities will be explored.

### *2.2. OpenAI model*

The chatbot's core intelligence is powered by the state-of-the-art OpenAI model [8-10]. OpenAI is renowned for its advanced natural language processing capabilities and has been a driving force in the development of conversational AI. The specific model utilized for this research is the OpenAI GPT-3.5 Turbo, which is a highly advanced and versatile language model.

DOI: 10.54254/2755-2721/46/20241435

The OpenAI GPT-3.5 Turbo model operates as the conversational engine of the chatbot. It enables users to interact with the system in natural language, making the user experience seamless and intuitive. Here's an overview of how the OpenAI model functions within the chatbot:

- Natural Language Understanding (NLU): The OpenAI model is exceptionally proficient at understanding the nuances of user queries. It can discern the intent and context of user inputs, allowing it to respond accurately to a wide range of questions and prompts related to Chinese universities.
- Response Generation (NLG): When a user interacts with the chatbot by entering queries or requests, the OpenAI model generates coherent and contextually relevant responses. These responses include detailed information about universities, advantageous disciplines, distinctive courses, historical backgrounds, and participation in national projects.
- Context Management: The OpenAI model excels at maintaining conversational context. It remembers previous interactions and user inputs, ensuring that the conversation flows naturally and coherently.

To integrate the OpenAI GPT-3.5 Turbo model into the chatbot, this study followed the setup process provided by OpenAI. This involved setting the environment variable "OPENAI_API_KEY" to securely access the model. The API key was obtained from OpenAI and is stored securely to ensure the confidentiality and integrity of the interactions with the model. The following code snippet demonstrates the setup process.

With the integration of the OpenAI model into the chatbot, it acquires the capability to deliver comprehensive and context-aware responses to user queries concerning Chinese universities. In the following sections of this paper, the focus will shift to examining the chatbot's performance and results, emphasizing its effectiveness in aiding users in accessing valuable information about Chinese universities. Please be mindful that the particular API key referenced in the code should be handled as confidential and stored securely in alignment with OpenAI's guidelines and best practices.

## 3. Results and discussion

In this section, this study presents illustrative examples of user interactions with the developed chatbot. These examples showcase the chatbot's capabilities and the nuances of its responses.

User Query 1:

User Query: "How is Anhui University of Traditional Chinese Medicine?"

Chatbot Response: "Anhui University of Traditional Chinese Medicine is a public university in Anhui Province, founded in 1959. It is a university under the Anhui Provincial Higher Education Revitalization Plan's 'Construction of Local Characteristic High-level Universities,' and also a pilot unit for the management of university staffing and filing system in Anhui Province. This study unable to provide detailed information about the specific situation and teaching quality of the university. You can obtain more accurate information by consulting the official website of the university or consulting relevant students or faculty."

User Query 2:

User Query: "What is said in the file?"

Chatbot Response: "The file mentions the following content: Anhui University of Traditional Chinese Medicine is a university under the Anhui Provincial Higher Education Revitalization Plan's 'Construction of Local Characteristic High-level Universities,' and is a pilot unit for the management of university staffing records in Anhui Province. Shanghai Lixin College of Accounting and Finance is a public full-time ordinary higher education institution in Shanghai, and is a high-level local applied university construction unit in Shanghai. Central South University of Finance, Economics, and Law is a key university directly under the Ministry of Education of the People's Republic of China. It is a key construction university under the national '211 Project' and '985 Project Advantage Discipline Innovation Platform' projects. Inner Mongolia University of Finance and Economics is a full-time ordinary higher education institution organized by the People's Government of Inner Mongolia Autonomous Region."

DOI: 10.54254/2755-2721/46/20241435

### *3.1. Discussion of results*

These examples showcase the chatbot's ability to provide informative responses to user queries regarding Chinese universities. In the first example, the chatbot accurately describes Anhui University of Traditional Chinese Medicine, including details about its founding, affiliations, and participation in regional education plans. It responsibly emphasizes the importance of referring to official sources and consulting relevant individuals for specific information.

The second example, while responding to a vague query, offers factual information related to multiple universities and their affiliations. This demonstrates the chatbot's capacity to provide comprehensive responses, even when faced with less precise user input. However, it highlights the significance of users formulating clear and specific questions to obtain the most relevant information efficiently.

These examples collectively underline the effectiveness of the chatbot in providing valuable information about Chinese universities based on user inquiries. As an aid for students navigating the complexities of university choices, the chatbot serves as a valuable resource, offering insights and guidance to facilitate informed decision-making in higher education.

## 4. Conclusion

In this work, initial steps have been taken towards the creation of a valuable educational tool. The chatbot's ability to deliver essential university information demonstrates its potential as a valuable resource for students navigating the complex landscape of higher education in China. Looking ahead, the next phase of the project will focus on implementing recommendation algorithms that consider users' individual profiles, including their academic achievements, grades, interests, and college's subject strengths. By leveraging machine learning and data analytics, the aim is to transform this chatbot into a comprehensive educational advisor capable of guiding students towards universities that align perfectly with their aspirations. This work represents just the beginning of a journey towards an intelligent and highly personalized university selection tool. With continued dedication and innovation, the full potential of this chatbot can be unlocked, making it an indispensable asset for students seeking the best-fitting educational institutions. In the future, the plan is to adapt and refine this chatbot to serve a broader range of educational decision-making tasks, further empowering students on their academic journeys.

## References

[1] Shanghai ranking 2023 https://www.shanghairanking.cn/
[2] Chang J 2006 Globalization and English in Chinese higher education World Englishes 25(3-4) 513-525
[3] Wan Y 2006 Expansion of Chinese higher education since 1998: Its causes and outcomes Asia Pacific Education Review 7 19-32
[4] Adamopoulou E and Moussiades L 2020 An overview of chatbot technology. In IFIP international conference on artificial intelligence applications and innovations (pp. 373-383) Springer Cham
[5] Ranoliya B R Raghuwanshi N and Singh S 2017 Chatbot for university related FAQs In 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI) (pp. 1525-1530) IEEE
[6] Luo B et al 2022 A critical review of state-of-the-art chatbot designs and applications. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 12(1), e1434.
[7] Du R 1990 Chinese higher education: A decade of reform and development (1978-1988) Brigham Young University
[8] Aydın Ö and Karaarslan E 2022 OpenAI ChatGPT generated literature review: Digital twin in healthcare Available at SSRN 4308687.
[9] Nandy A and Biswas M 2017 Reinforcement Learning: With Open AI, TensorFlow and Keras Using Python. Apress.
[10] Lee J S and Hsiang J 2020 Patent claim generation by fine-tuning OpenAI GPT-2 World Patent Information 62 101983