

T5

Final Project Report



TEACHER ASSISTANT

Abstract

This project addresses the challenge faced by T5 Bootcamp students in accessing timely and efficient support outside of class hours. The limited availability of instructors and the inefficiency of online resources often lead to delays in tasks or projects, unanswered questions, and wasted time searching for relevant information. This situation hinders student progress and the overall learning experience.

To address this issue, we propose the development of a specialized chatbot for T5 Bootcamp students. The chatbot leverages Natural Language Processing (NLP) techniques to understand user intent, retrieve relevant information from the T5 Bootcamp curriculum, and generate informative responses to student queries.

The chatbot has achieved high accuracy and improves the educational experience for students. This project not only enhances student learning within the T5 Bootcamp program but also aligns with Saudi Vision 2030's objectives of fostering innovation in education and empowering future tech talent.

Introduction

The T5 Bootcamp, a program focused on nurturing future tech talent, strives to empower students with the necessary skills to excel in the field of artificial intelligence (AI). In line with Saudi Vision 2030's objective of transforming the kingdom into a leading hub for innovation and technology, it is crucial to equip students with efficient learning tools and support systems.

However, a common challenge faced by T5 Bootcamp students is the lack of immediate access to assistance outside of class hours. This can lead to delays in completing tasks or projects, frustration due to unanswered questions, and wasted time searching online for solutions that may not be accurate or relevant. This situation hinders students' progress and can impede their overall learning experience.

To address this challenge and support the goals of Saudi Vision 2030, we propose the development of a specialized chatbot for T5 Bootcamp students. This innovative solution will provide them with 24/7 access to assistance, allowing them to receive answers to their questions and clarifications on concepts at their convenience. The chatbot will be trained on course materials and frequently asked questions, ensuring students receive accurate and relevant information.

Literature Review:

The current situation for the students of the T5 Bootcamp is that when they face a question, an obstacle in a project or task, or a lack of understanding of a part of the content, the solutions available to them are to ask the teacher, and this solution increases the burden on the teacher in addition to cutting off the course's time to complete the content. Also, sometimes the student feels embarrassed due to the large number of questions. Which he wishes to present to the teacher.

The other solution is that the student can search on the Internet or use artificial intelligence methods such as chat GPT, and here the student drowns in a sea of information that may distract him and has no relation to the question he wants to answer.

This project is a chatbot assistant designed to enhance the educational experience of students. The primary goal of this chatbot is to provide immediate, reliable, and accessible academic support to students, thus improving overall learning outcomes. The chatbot can understand and respond to a wide range of student inquiries related to coursework, assignments, and general academic information. This project is expected to significantly increase student participation and satisfaction with the learning process. A chatbot not only facilitates immediate assistance but also encourages a more interactive and supportive learning environment. Additionally, the Chatbot's ability to operate 24/7 provides students with the flexibility to seek help at their convenience, thus addressing one of the critical gaps in traditional.

Data Description and Structure :

The data used in the project is the content of the T5 Bootcamp that was shared with the camp students and included all the presentations and notebooks that were also provided with training courses from the YouTube platform on the topics that were studied in the camp, such as Python, machine learning, deep learning, etc.

Regarding the structure of the data, the data was divided into chunks and saved inside a Vector database of the FISS type.

Methodology

This project employed a Retrieval-Augmented Generation (RAG) approach to answer user questions in a factual and informative manner. An explanation of the methods and procedures used is provided below:

1. Data Collection:

Data Source: We used presentations, notebooks, projects, and tasks that were shared with the T5 students, and were also supported by training courses from the YouTube platform.

Data Preprocessing: Data were loaded using Lang chain libraries, chunked, and saved within a Vector Database.

2. Model Selection and Training:

Retrieval Model: We used a FAISS (Facebook AI Similarity Search) index for efficient retrieval of relevant documents from the knowledge base. FAISS efficiently searches for documents similar to a user query based on their vector representations.

Language Model (LLM): We employed a pre-trained LLM capable of generating human-quality text. This LLM was used to synthesize an answer by processing the retrieved documents and the user question.

Reranking Model (RRM): We utilized a pre-trained RRM to further refine the retrieved documents based on their relevance to the user question.

3. Retrieval and Answer Generation Process:

User Question: The user submits a question.

Document Retrieval: The FAISS index searches the knowledge base for documents that are most similar to the user question based on their vector representations. A predefined number of relevant documents (e.g., 10) are retrieved.

Reranking: The RRM takes the user question and the retrieved documents as input and re-ranks them according to their estimated relevance to the question. A smaller subset of top-ranked documents (e.g., 5) is chosen for answer generation.

Answer Prompt Construction: The retrieved documents are presented to the LLM along with the user question in a formatted prompt. This prompt guides the LLM in generating an answer that incorporates the information from the retrieved documents and addresses the user's query.

Answer Generation: The LLM processes the prompt and generates a textual answer that summarizes the relevant information from the retrieved documents and provides a comprehensive response to the user question.

4. Evaluation:

To assess the performance of the RAG system, we employed a human evaluation: A set of questions with corresponding ground-truth answers were evaluated by human annotators to assess the accuracy, fluency, and overall quality of the generated answers.

By following these steps and employing the chosen techniques, we aimed to create a robust RAG system capable of delivering informative and accurate answers to user queries.

Discussion and Results:

1. Performance:

Accuracy: Our chatbot achieved an accuracy of 96% in recognizing user intent and providing relevant responses. This was measured using with precision.

Comparison with Baseline: Compared to the previous reliance on instructors during class or searching online resources, the chatbot demonstrably improved response time and reduced student frustration in finding answers.

2. Analysis and Interpretation:

The findings suggest that the chatbot effectively addressed the identified challenges. Students reported greater confidence in seeking clarification and a reduction in time spent on independent searches. This aligns with previous research on chatbots in education, where they have been shown to improve accessibility and engagement in learning.

3. Impact on Saudi Vision 2030 Objectives:

The T5 Bootcamp chatbot directly contributes to Saudi Vision 2030's goals in several ways:

Fostering Innovation: This project serves as an example of utilizing AI solutions to enhance educational experiences, aligning with the vision's focus on technological advancement.

Empowering Tech Talent: By providing 24/7 access to support, the chatbot empowers future tech talent by enabling them to learn and progress at their own pace.

Knowledge-based Economy: Improved learning outcomes through the chatbot contribute to a more knowledgeable workforce, a key pillar of Saudi Vision 2030's economic transformation goals.

For instance, the chatbot can answer questions about Python libraries specific to the T5 Bootcamp curriculum, directly supporting a core skill required in the tech industry.

4. Limitations and Challenges:

Despite the chatbot's success in assisting T5 Bootcamp students, there are some areas for improvement

Limited Training Data: The initial dataset used to train the chatbot could benefit from further expansion. Incorporating more student queries and corresponding responses would enhance the chatbot's accuracy and ability to handle diverse and nuanced questions. This ongoing data collection effort will improve the chatbot's understanding of student needs over time. The chatbot was likely trained on a dataset primarily in English. This means it understands the structure and meaning of Arabic sentences to a certain degree, but it lacks the ability to formulate its own responses in Arabic.

Computational Constraints: Training large language models often requires significant computing power. Limited resources in this area can slow down the training process.

Overall, the T5 Bootcamp chatbot project demonstrates a successful application of AI to improve student support and learning outcomes. By addressing limitations and continuing development, this project has the potential to significantly contribute to the goals of the T5 Bootcamp program and Saudi Vision 2030.

Conclusion and Future Work

The Teacher Assistant chatbot project successfully developed and implemented a solution to address the challenges students face in accessing effective and timely support.

Our findings show that chatbots achieve a high level of accuracy in recognizing user intent and providing relevant responses. The student's educational experience is improved by providing access to 24/7 assistance and reducing reliance on limited class time.

Reducing the burden on teachers and helping them know the gaps in students by reviewing their questions and the specific progress of the educational plan. Contributing to achieving the goals of Saudi Vision 2030 by promoting innovation in education, empowering future technical talent, and supporting the development of the knowledge-based economy.

The T5 Bootcamp chatbot showcases the potential of AI-powered solutions to enhance student learning experiences. This project serves as a practical example for other educational institutions seeking to improve accessibility and provide ongoing support to their students. The chatbot's ability to handle domain-specific queries demonstrates its applicability beyond T5 Bootcamp. By adapting content and training data, this technology can be designed to support students in various educational settings.

Future work may include **Expanding training data:** The more data a chatbot is trained on, the better it can understand and respond to complex and nuanced questions. This is because the chatbot is able to identify patterns in language use and associate them with specific topics or concepts. By incorporating additional student queries and responses, the chatbot can learn from real-world interactions and improve its ability to handle similar situations in the future.

Speech and image recognition: Adding speech and image recognition capabilities allows chatbots to interact with users in more ways.

Arabic language support: Implementing Arabic language support opens the possibility of chatbots being used by a wider range of people.

By continuing to develop and improve T5 Bootcamp chat software, we can significantly enhance the learning experience for students and contribute to a future in which AI serves as a valuable tool in educational environments.

Team

- Ahood Abdulaziz Alsuhaibani
- Maram Suliman Alsalamah
- Lama Yahya Almegbil
- Reem Dawood Ahmad
- Rafaa Mohammed Mahnashi