

AN ALTERNATIVE APPLICATION TO CHATGPT THAT USES RELIABLE SOURCES TO ENHANCE THE LEARNING PROCESS

Mirela ȘORECĂU, Emil ȘORECĂU

Technical University of Cluj-Napoca, Romania,
“Nicolae Bălcescu” Land Forces Academy, Sibiu, Romania
mirela.doicin@gmail.com

Abstract: ChatGPT is a type of language model that uses machine learning algorithms to generate responses to natural language input. Its advanced technology can revolutionize the way humans interact with machines by creating more natural and intuitive communication. However, the accuracy of ChatGPT's responses may be compromised by biased or inaccurate data, which highlights the importance of carefully evaluating its output. Furthermore, early versions of ChatGPT were found to produce academic papers with missing references, which may compromise the credibility of research in the academic publishing industry. This underscores the need to establish regulations and guidelines that ensure the ethical and transparent use of these technologies, and to avoid relying solely on automated tools like ChatGPT without thoroughly reviewing the literature. In order to minimise the potential for inaccurate information in the classroom, an intuitive application based on the AI language (OpenAI - also used by ChatGPT) has been implemented. The application has been designed in such a way that it does not use internet databases as a search source and the answers to the questions are based on pre-established documentation that has been checked/certified by the educational institution/professors. The developed application can be installed either locally on a personal computer or on a file server/online library etc. and does not require access to a reliable Internet connection as in the case of ChatGPT.

Keywords: ChatGPT, Artificial Intelligence, OpenAI, education, chatbot

1. Introduction

The impact of artificial intelligence in education is significant, as it can enhance personalized learning experiences, improve teaching efficiency and support research and development. However, ethical and technological challenges need to be considered and addressed, and the impact on teaching and learning should be monitored and evaluated as AI continues to develop.

ChatGPT is an innovative tool that has the power to revolutionize the way we interact

with technology. While it can be a valuable resource, it's important to recognize that its results are not always accurate or reliable, especially if the sources it uses are biased or outdated. As an artificial intelligence language model, ChatGPT has no natural understanding of the correctness or quality of the sources it uses. The quality of the sources it uses depends on the training data used to train the model [1]. There are also concerns about the ethical and transparent use of ChatGPT and similar technologies, and the potential consequences of relying

solely on automated tools such as ChatGPT for tasks that require human judgement and critical thinking. It is important to carefully assess the accuracy and reliability of ChatGPT and to develop new guidelines and regulatory frameworks to ensure that these technologies are used ethically and responsibly. Based on this premise, it is important to develop reliable and accurate technology, using only trusted sources that have undergone a scientific evaluation process, and taking into account ethical issues.

Recent publications highlight the problem of missing references in academic research papers generated by ChatGPT, which could raise concerns about the integrity and credibility of research without adequate references. A problem we have experienced, which is also mentioned in this paper [2], is that ChatGPT provides fictitious citations to non-existent articles, which can cause additional problems for students regarding the veracity of the information, but also for the publication of academic papers. There are already several studies on the impact of using ChatGPT on specific fields of research: education, economics and finance, medicine, programming, cybersecurity etc. In [3], the ability of natural language processing (NLP) to improve academic performance in economics and finance using ChatGPT is investigated. The study concludes that ChatGPT has the potential to help researchers with data analysis, scenario generation and communication of results. However, there are limitations to using chatbots or similar tools in research, including dependence on data quality and diversity, lack of domain expertise, and ethical considerations.

Another study [4] evaluates the performance of ChatGPT as a scheduling assistant and finds that it is effective at solving common scheduling problems, but struggles to solve new problems and has a limited attention span. However, it can provide valuable information about the

programmer's intentions and can be useful in testing research. The paper [5] also highlights the language model's ability to perform various programming-related tasks, provide guidance for validating and correcting code, and answer technical questions, making it a powerful and versatile tool for the programming industry. A study in the medical field explored the potential of using ChatGPT to simplify radiology reports for patients [6]. The study concluded that the quality of the simplified reports generated by ChatGPT was factually accurate, but also identified errors that could be potentially harmful if medical staff were not involved in the process.

The use of artificial intelligence, and in particular linguistic models such as ChatGPT, to improve academic performance is a growing area of interest and research [7], [8], [9]. While there is certainly potential for these technologies to provide benefits in this context, it is important to ensure that the data provided by these models is accurate and reliable. This is particularly important in fields where accuracy is essential, such as medicine or engineering. By analyzing the accuracy of the data provided by ChatGPT and similar models, researchers can determine the potential impact of these tools on academic performance [10] and identify areas where improvements are needed.

Although currently the primary input used by ChatGPT is in the form of text, there are ongoing initiatives to integrate other forms of media content such as images, video, audio, etc. into the functions of the model. This approach makes ChatGPT much more responsive to multi-format input, providing detailed and comprehensive solutions for a variety of applications [11]. The ChatGPT could also be improved on the basis of data specific to each field of research/specialization. This is desirable as it would achieve significant performance for individual applications, but it also becomes a challenge due to the insufficient

availability of tagged information. However, we believe that this process of approaching the application for particular domains can bring significant improvements in the accuracy and relevance of the answers provided.

The potential and limitations of using ChatGPT in academic research are equally obvious. The most serious issue related to its capabilities concerns the erroneous formulations and references generated by ChatGPT [12].

An alternative solution to ChatGPT, which we propose in this paper, is to use only trusted sources by implementing a filtering system that allows access only to verified and pre-established sources, such as books, scientific articles and textbooks that have been peer-reviewed and certified.

This complementary ChatGPT approach of implementing a filtering system and relying heavily on trusted sources can be used in applications where accuracy and reliability of information is essential. In areas such as scientific research, academic work, medical research/diagnostics, legal research, etc., access to less accurate or biased information can have major implications. Most importantly, such a system can provide a remarkable benefit to the educational environment by encouraging students to access only trusted sources. In this way, it encourages the use of reliable information, promotes academic ethics and integrity and, secondarily, enhances the value of student work. In addition, this application can be integrated into virtual libraries to ensure that students have access to quality resources that have been reviewed by experts. Based on a filtering system, it can be seen as an effective tool to promote more sophisticated research methods and encourage students to consider authentic, accurate and reliable information. A quick search system containing a database of accurate information can improve the efficiency of student learning by reducing time, ensuring accuracy and accessibility, providing targeted results,

encouraging critical thinking, and supporting personalized learning.

2. Materials and Methods

The aim of this work was to implement a custom application that would allow the use of artificial intelligence resources used by OpenAI to provide information from local and verified sources.

The Python programming language was used for this purpose. Python programming is a versatile and powerful tool for developers of all levels. The language's clean syntax and ease of use make it an excellent choice for beginners, while its extensive libraries and frameworks make it a popular choice for experienced developers. Python is widely used in various fields including web development, scientific computing, information science data analysis, artificial intelligence, machine learning and beyond [13]. In addition, Python's cross-platform compatibility allows it to be used on multiple operating systems, making it an ideal choice for software development. Langchain was also used to integrate OpenAI into the application. LangChain is a powerful framework that simplifies the development of applications using Large Language Models (LLMs). It provides a single, customizable interface capable of managing different LLMs, including prompt management, chaining, data augmentation, agent orchestration, storage and evaluation. This versatile framework allows developers to seamlessly integrate LLMs with their real-world workflows and data with minimal effort.

LangChain's advanced features enable LLMs to solve real-world problems more effectively by integrating external data sources and orchestrating interactions with multiple components in a specific sequence. Another library that was used was ChromaDB. ChromaDB is a schema-free vector database designed specifically for use in artificial intelligence applications. It is both lightweight and very powerful,

allowing efficient storage, retrieval and management of vector data (embeddings), which is essential for LangChain and OpenAI-based document chat applications. By using ChromaDB, it is possible to easily manage vector data to improve the performance and functionality of the AI applications.

```
1 from langchain.embeddings.openai import OpenAIEmbeddings
2 from langchain.vectorstores import Chroma
3 from langchain.text_splitter import CharacterTextSplitter
4 from langchain import OpenAI, VectorDBQA
5 from langchain.document_loaders import DirectoryLoader
6 import os
7 import tkinter
8 from tkinter import filedialog
```

Figure 1: Importing the required libraries

The main steps of the application are as follows:

1. Create a local folder and copy the documents that will form the basis of the AI search, and implement in Python the lines of code needed to load the data from this location.

```
14 loader = DirectoryLoader(location + "/", glob='**/*.doc')
15 docs = loader.load()
```

Figure 2: Load local documents

2. Split the text found in the files into chunk_size of 1000 with a chunk_overlap of 20 to facilitate searching and generate embeddings using OpenAI models.

```
16 char_text_splitter = CharacterTextSplitter(chunk_size=1000, chunk_overlap=20)
17 doc_texts = char_text_splitter.split_documents(docs)
```

Figure 3: Split text into small chunks

3. The final step is to create and maintain a Chroma database containing the embeddings and associated metadata. This allows efficient storage and retrieval of document embeddings for text analysis and AI-based interactions.

```
21 docsearch = Chroma.from_documents(doc_texts, embeddings)
22
23 model = VectorDBQA.from_chain_type(llm=OpenAI(), chain_type="stuff", k=2, vectorstore=docsearch)
```

Figure 4: Create and maintain a Chroma database

The graphical user interface of the application was configured using Tkinter. Tkinter is a Python GUI toolkit for building desktop applications with an interactive graphical user interface. It is pre-installed as a standard library with most Python

distributions and provides a comprehensive set of tools and widgets that simplify the process of designing a desktop application. It is based on the Tcl/Tk GUI toolkit and provides an easy-to-use interface that allows Python developers to easily create and customize their GUI applications.

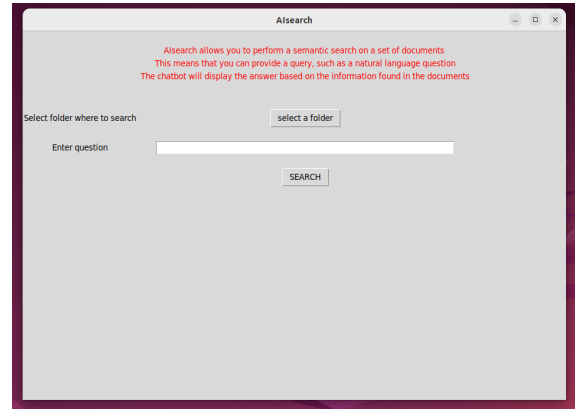


Figure 5: Graphical user interface of the application

3. Results and Discussions

To test the application, a 'Documents' folder was created containing 5 documents, generically named 'Book1-Book5'. In these documents, information related to the “Radio Communications” courses has been added. These will be the basis for the creation of the Chroma vector database. Using the graphical interface, this folder was selected.

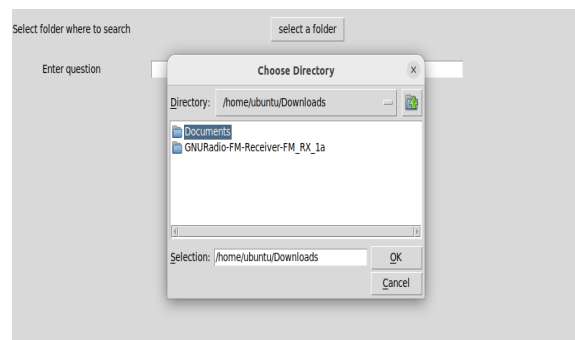


Figure 6: Select AIssearch source folder

The application is then ready to receive questions and answer them based on the source documents. Once the question is asked, the answer is returned along with the source information of the text fragment from which the required information was taken. This allows us to ask any question

related to the text document using the OpenAI language model in the background.

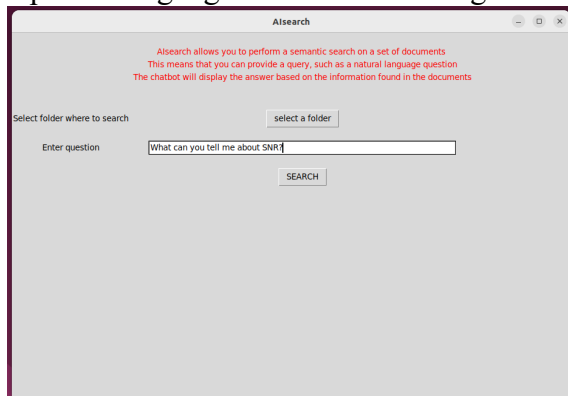


Figure 7: Ask question using Alsearch

After entering the question, there are two possibilities: one where the information is found in the local documents and the answer is generated based on it, or another where there is no information to formulate an answer.

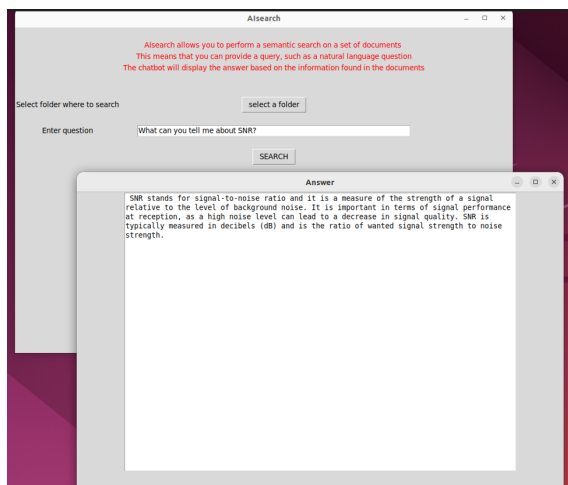


Figure 8: Getting the answer to the question

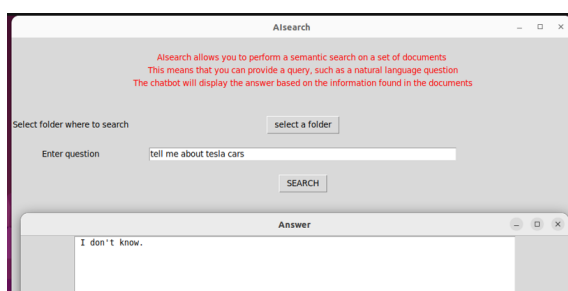


Figure 9: Receive "I don't know" message if there is no information to answer

The application also provides information that shows the documents and source text on which the answer is based, allowing users to verify the accuracy and authenticity of the answer.

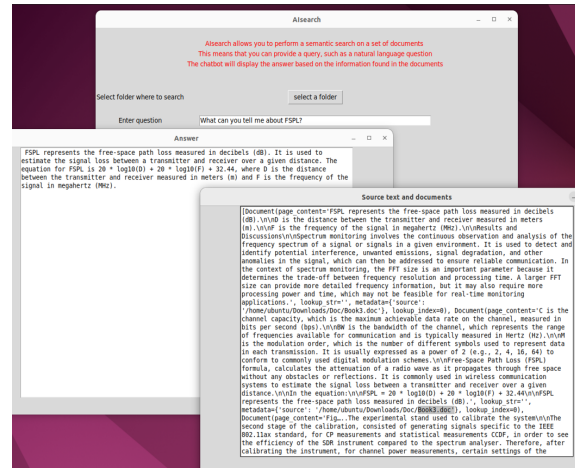


Figure 10: Receive answer with source information

Another feature of the application is that it answers questions in the language in which the question was asked, even if the documents are in a language other than the language in which the question was asked. In the example below, a question was asked in Romanian and answered in Romanian, even though all the source documents were in English.

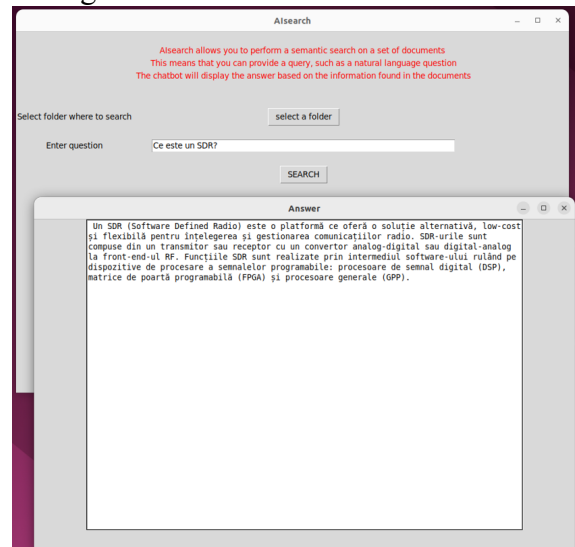


Figure 11: Query and response in a language different from the source documents

4. Conclusions

Implementing such an application that provides fast and reliable answers using filtered sources in the educational environment offers several benefits. It promotes performance and productivity, saving students time and allowing them to cover more material. It increases accuracy

and credibility by providing reliable sources of information, reducing the risk of misinformation and promote critical thinking and research skills as students evaluate and question information from other sources.

A quick search system with a database of accurate information may be more effective for students who need to find reliable and relevant resources quickly and conveniently. A system that only targets information in a specific field may be more effective for students studying that field, but may not be as versatile or comprehensive as a broader search system. ChatGPT is a broad search system that can also be a useful tool for students, but may not always provide accurate or relevant information, especially in specialized or technical areas.

The choice between a targeted and a broader search system depends on the specific needs and requirements of the users, but it should be noted that ChatGPT, although impressive, cannot replace human intelligence.

Such an application as "AISearch" can be developed and extended in multiple domains, involving a broader process of labeling and classifying data based on established domains, creating a well-selected dataset, training specialized models, leveraging existing knowledge, diversifying data, and evaluating performance. All of these methods lead to high levels of accuracy for different research domains, but require domain expertise and a sustained effort of continuous development.

References List

- [1] Haque M.A. A Brief analysis of "ChatGPT" – A revolutionary tool designed by OpenAI. EAI Endorsed Transactions on AI and Robotics. 2023.
- [2] Lund B.D., Wang T., Mannuru N.R., Nie B., Shimray S.R. & Wang Z. ChatGPT and a new academic reality: Artificial Intelligence-written research papers and the ethics of the large language models in scholarly publishing. *Journal of the Association for Information Science and Technology*. 2023; 74: 570 - 581.
- [3] Alshater M.M. Exploring the role of artificial intelligence in enhancing academic performance: a case study of ChatGPT. *SSRN Electronic Journal*. 2022.
- [4] Tian H., Lu W., Li T.O., Tang X., Cheung S.C., Klein J. & Bissyand'e T.F. Is ChatGPT the ultimate programming assistant - How far is it? In *ArXi*. 2023; abs/2304.11938.
- [5] Biswas S. Role of ChatGPT in Computer Programming. *Mesopotamian Journal of Computer Science*. 2022.
- [6] Jeblick K., Schachtner B.M., Dext J., Mittermeier A., Stüber A.T., Topalis J., Weber T., Wesp P., Sabel B.O., Ricke, J. & Ingrisich M. ChatGPT makes medicine easy to swallow: an exploratory case study on simplified radiology reports. In *ArXiv*. 2022; abs/2212.14882.
- [7] Lund B.D. & Ting W. Chatting about ChatGPT: How may AI and GPT impact academia and libraries? *SSRN Electronic Journal*. 2023.
- [8] Zhai X. ChatGPT user experience: implications for education. *SSRN Electronic Journal*. 2022.
- [9] Ratnam M., Sharma B. & Tomer A. ChatGPT: Educational Artificial Intelligence. *International Journal of Advanced Trends in Computer Science and Engineering*. 2023; 12(20): 84.91.
- [10] Qasem F. ChatGPT in scientific and academic research: future fears and reassurances. *Library Hi Tech News*. 2023.

- [11] Hariri W. Unlocking the Potential of ChatGPT: A comprehensive exploration of its applications, advantages, limitations, and future directions in Natural Language Processing. In ArXiv. 2023; abs/2304.02017.
- [12] Rahman M., Terano H.J., Rahman N., Salamzadeh A. & Rahaman S. ChatGPT and Academic Research: A Review and Recommendations Based on Practical Examples. Journal of Education, Management and Development Studies. 2023.
- [13] Raschka S., Patterson J. & Nolet C.J. Machine Learning in Python: Main developments and technology trends in data science, machine learning, and artificial intelligence. ArXiv. 2020; abs/2002.04803. <https://doi.org/10.3390/info11040193>.