



UN VOLUNTEERS

UN VOLUNTEER DESCRIPTION OF ASSIGNMENT

Id 1722716229608704

VMAM Id 1722716229608704

Opportunity Type (Online/Onsite)? Online

Opportunity Title Develop app to query PDF documents with Large Language Model locally

UN VOLUNTEER DESCRIPTION OF ASSIGNMENT

Host Entity	Bread for the World
Country of Assignment	Germany
Number of UN Volunteers	1
Sustainable Development Goal	2. Zero hunger
How many hours per week will the volunteer be needed?	11 - 15
Duration	2

DOA Details

Organisation mission and objectives

Brot für die Welt (Bread for the World) is the globally active development and relief agency of the Protestant Churches in Germany. In more than 85 countries all across the globe we empower the poor and marginalised to improve their living conditions. Key issues of our work are food security, the promotion of health and education, the access to water, the strengthening of democracy, respecting human rights, keeping peace and the integrity of creation. Brot für die Welt takes steps to make sure that there is enough food for everyone. Because fighting hunger becomes increasingly important in times of climate change and ever scarcer resources. The work of Brot für die Welt primarily focuses on the support of projects in the countries of the Global South. An essential feature of our projects is the close and continuous cooperation with local, often church-related partner organisations. Upon request, Brot für die Welt provides them with specialists and volunteers. Through lobbying, public relations and education in Germany and Europe we seek to influence political decisions in favour of the poor and to raise awareness for the necessity of a sustainable way of life.

Assignment context

Since the release of ChatGPT, numerous third-party apps such as <https://myaskai.com> or <https://askyourpdf.com> have been released that allow users to query PDF documents with ChatGPT. While useful, these apps require the upload of documents onto third-party servers, as well as the submission of contents to OpenAI's servers, both of which is not always an option for Brot für die Welt due to data security and privacy concerns. Our organization therefore needs an in-house app that provides similar functionality to staff without sending internal data to third parties. Having such a tool in-house will allow staff in the organization to process large documents more effectively saving time and resources in the process.

Task description

Develop a self-hosted PDF indexing and querying app for Brot für die Welt that uses an open-source Large Language Model (LLM). The app should allow a user to upload a PDF file, index its contents for keyword searching, and query the indexed text via a chatbot interface. The app should send the portions of the PDFs that contain the keywords to the language model who will formulate a response to be sent back to the user. The app should run locally on Brot für die Welt's servers without sharing data with third-party providers. The volunteer should provide source code and documentation upon completion of the project.

DOA Requirements**DOA Requirements****Required education level****Area(s) of specialisation****Required experience****Required skills and experience**

- Strong experience in developing web applications and software solutions, including front-end and back-end development.
- Proficiency in programming languages such as Python, JavaScript, and HTML/CSS.
- Familiarity with chatbot development and natural language processing.
- Familiarity with open source language models as well as the GPT API.
- Experience in integrating open-source libraries and APIs for PDF processing and indexing.
- Understanding of database design and management, as well as server deployment.
- Knowledge of software security and best practices in data privacy and protection.
- Excellent problem-solving skills, attention to detail, and ability to work independently.

Language**Language Language skills Language requirement****Area of Expertise****What type of task do you need support with?****Does this project include an additional, secondary type of task?**