Extracting Green Knowledge for the Land of Wood and Water

Impressed by the launch of NVIDIA AI Workbench, followed up by seeing the posted competition regarding taking a level 1 provided retrieval-augmented generation project and making use of its limitless potential to create a program of interest, this Jamaican-born team’s excitement to build blossomed once more. Upon properly reading through the instructions, we quickly got to work.

After cloning the forked repository through following amazingly illustrated and written README of the repo, our primary focus was comprehending the default setup which wasn’t difficult since the developers had done a great job in presenting the code in very clean and easy to understood outline. After skimming through the documentation, clarifying the flow of the program, and gathering a basic understanding of the great power we now possess, responsibly choosing how to use it had become our next step. Through brainstorming on how we can uniquely revolutionize the use of an opensource text generative program with an easily customizable database, we had quickly obtained the idea of allowing for its use to further assist the golden egg laying goose of our country’s economy – Jamaican agriculture. Our idea was simple yet powerful; to leverage the local agricultural and environmental data to create an AI-powered assistant that helps farmers and policymakers. We called it AI-griculture JM. With this main task in mind, we had planned to engrave in it the following features:

1. Document Housekeeping: To store a wide range of publicly available documents regarding Jamaica’s political and environmental conditions such as government agricultural policy documents, research papers on soil health and maintenance, weather forecasts and historical climate data, local market trends, and export reports for crops.
2. Chat Interface: To communicate the stored data through natural language prompts (e.g. “Give some strategies for organic yam farming” or “How will the upcoming drought affect coffee production?”)
3. Local Data Integration: To incorporate specific data relevant to Jamaican agriculture, such as soil conditions, rainfall patters, and pest outbreaks.
4. Recommendation System: To, without need for user prompting, suggest optimal farming operations relevant to the present events (e.g. inform on optimal planting dates, crop rotations, or irrigation schedules)
5. Offline Capabilities: To provide a various amount of the previously stated features under the circumstances of the user lacking access to the internet.

After reviewing the template and doing additional research, the ideas of how we would begin tweaking the program to go in line with our ideas became apparent as we began the program’s development:

1. **Refurbish the front-end**: We followed up on the initial use of Gradio by the repo to create a user-friendly interface for rural inhabitants. Focusing on functionality and simplicity, we ignored extensive graphics, while maintaining a visually appealing design. The goal was to make the interface accessible and adaptable to rural inhabitants' accommodation to modern technology. Since our program was going to replicate a basic one-on-one conversation between man and machine, we knew that one input textbox for the user’s prompts followed by the submit button and a display for the model’s response would be the ideal layout for the average farmer.
2. **Optimize the database**: Of course, if we are going to make use of a model that pulls data from a database… we’ll need a database that has data. Luckily, we were provided with a prebuilt subpar vector database, which would serve as more than enough for our goals. Given our plan to build the model’s knowledge mostly around agricultural production, we simply needed to search for text data relevant to the concept, then embed them in a form that the program will accept. Such data wasn’t hard to find, as anything dealing with sustaining human lives has been heavily researched and outsourced. We collected multiple pdf files and inserted each into our vector database, which pretty much laid the initial groundwork of our project but introduced an easily solvable problem.
3. **Optimize the retrieval method**: Although skewering through every square inch of the internet and reloading all news outlets manually would be a glorious task, unfortunately, the mindset of a programmer plagued us to do otherwise. One of the easiest to brag about features of a hybrid rag would be its ability to update itself through a live data source. Therefore, automation was inevitably going to be a massive part of our data gathering, so we wasted no time in getting to work. We decided on two main ways of lively gathering data: API requests and web scraping. After exploring many generic agricultural and atmospheric data based APIs we could leverage to our country’s benefit, we decided on using Weather Stack and News API. We implemented methods to pull and store weather data for cities in Jamaica, as well as any news data relating to agriculture. They were stored in json files then formatted into text files for embedding.
4. **Optimize in use models**: Arguably the most crucial part of this project was going to end up being the entity that makes it work. Using a properly tuned model built to apply our data gathering was going to be just as important as deciding on said data gathering. We noticed that we could simply edit the hyperparameters of the given models, but we also noticed we were better off not using those models. Concluding that an outside-built LLM would better get the job done, we installed Gemini into the mix after we fine-tuned it for our specific goals.
5. **Optimize hallucination**: So apparently LLMs, much like the teacher’s pet, have a tendency to experience psychosis when asked a question they don’t know the answer to. A benefit of implementing a RAG system with them would be the provided solution to such an issue by telling them it is okay to say, “I don’t know.” Naturally, as this feature is left to the discretion of the insentient clump of algorithms, helping the model come to terms with this reality is going to require some form of fine tuning. This would be considered the mitigation of the hallucination of the model and, overall, our last needed task. We would do this by…

All in the end, the completion and success of our project should in multiple ways benefit the following people of our nation:

* **Farmers**: The program is capable of providing accessible and tailored advice for optimizing crop yields, thus applying better calculated strategies for handling the effects of climate change, enhancing Jamaica’s food security.
* **Policymakers**: The program can offer data-driven insights to develop more effective agricultural policies and strategies for sustainability.
* **Researchers**: The program enables easy retrieval of information across vast documents for faster research and knowledge-sharing.

From the beginning to the end of this project, we overall enjoyed the experience of building something of great potential in value to our country, as well as learning many new model tuning techniques along the way. Although a fairly simple process, the development of this application will leave a mark an everlasting mark on all our skills and experience as we can only consider this the beginning of a multitude of NVIDIA AI Workbench assisting projects to come.