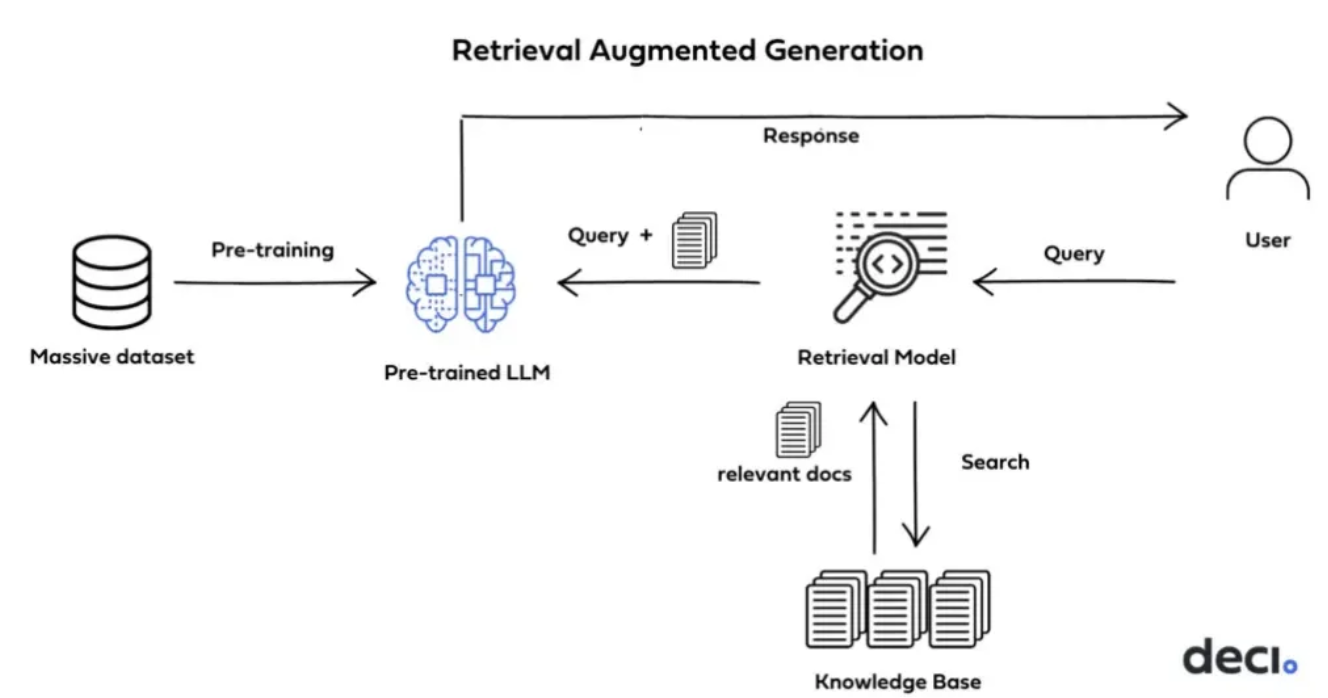
### **Group members:** Youzhang He, Yixiao Ling

1. **Project Summary:**

Build a chatbot providing psychological-health suggestions given the questions from users with large language models (LLM) and the Langchain retrieval-augmented generation (RAG) system.

1. **Project Details:**

Here is the general pipeline of the chatbot:



<https://deci.ai/blog/fine-tuning-peft-prompt-engineering-and-rag-which-one-is-right-for-you/>

We will download the psychological health answer and query dataset containing question and answer pairs and store it in the vector database after tokenizing the sentences in the dataset and applying transformers to embed the tokenized sentences into vectors.

Then we will set up the pipeline of the retrieval model provided by Langchain, LLM model (temporarily we choose GPT 3.5, but we may explore other models).

When the user inputs a question as a query, the transformer will transform the query into vectors, and the retrieval model will match corresponding texts from the psychological health answer and query dataset stored in the vector database. The pipeline will combine the query and the matched text and transfer them to GPT 3.5 model, the GPT 3.5 model will generate answers based on the query and the matched texts.

Moreover, we will set a chat history buffer to store the most recent chat history since we want the chat to be more consistent. The chat history buffer will automatically drop oldest chat history after running out of storage memory.

1. **Advanced Python:**

We need to use Python libraries and packages including: Langchain, OpenAI, pandas, and numpy.

1. **Data source or API:**

The psychological health answer and query dataset can be accessed through this link: <https://www.kaggle.com/datasets/thedevastator/mental-health-chatbot-pairs>

And we need the OpenAI API key for GPT 3.5, which is free.

1. **Willing to share:**

We are willing to share our projects.