AI ENGINEER ASSESSMENT

Scenario

Assuming that our company is developing an advanced AI-powered assistant to enhance knowledge retrieval / news retrieval for enterprise users. The assistant should help users by answering questions based on two primary sources:

1. **Internal Knowledge Base** – The assistant should retrieve relevant information from an internal csv dataset containing news articles and their sources and return this information to the user. The usage of a naïve method[1] is *not encouraged*. Instead, you should use Retrieval-Augmented Generation to answer the user.
2. **Web Search** – If the internal knowledge base does not provide a sufficient answer, the assistant should fetch real-time data from the web using a search engine API and return this web-searched information to the user as well with a note saying that this is searched from the web.

The 'test questions' that we will use will be as follows:

1. What are some initiatives launched by MCMC?
2. Adakah SSM terbabit dengan kes-kes mahkamah?
3. What is the status of the Malaysian economy in 2025 and associated headwinds?
4. *Additional questions we will ask later.*

The information returned should clearly indicate what is the source of the information (from which news source if it is using the Internal Knowledge Base or mention it is web search if the information is from web search). You should also suggest all the ways to a) evaluate and b) improve the accuracy and consistency of the responses.

Once you are done, add another functionality to your application such that the user can ask a question regarding the statistics of the dataset (such as 'how many positive and negative news are there' or 'how many of the news are before June 2025'. This may be in the same chat window as the core functionality (more complicated) or a separate chat interface (simpler). Again, you should not be using naïve[1] methods.

[1] A naïve method involves sending the contents of the entire csv to an LLM and ask it to provide answers to the user's question based on said content. This is not practical in real life applications.

Dataset

You are provided with a dataset (news.csv) containing multiple columns but the most important ones are the **Title, Article Content, Summary, Author** and **Sentiment** fields. Utilize this dataset as the knowledge base for your application.

Title - Title of the news article
Article Content - Content of news article
Summary – Summary of the content
Author – The 'source' of the news/news source. It can be a newspaper or a person.
Sentiment – This indicates whether it is a positive news or negative news.

Tips: You are encouraged to integrate **Gemini's free API** for enhanced responses. Open-source models are also welcomed.

Deliverables

Candidates should submit the following:

1. **A functional web application** which allows the user to ask questions on news from the internal knowledge base or the web as well as statistics on the news in the knowledge base and have the AI-powered assistant answer the user. You may use frameworks like Streamlit for rapid front-end development.
2. **Source code repository** (GitHub) with clear documentation.
3. Provide a README file with a concise overview of the web application, including its functionality, setup instructions, and key components.

Evaluation Criteria

- **Functionality** - Assess the accuracy and effectiveness of the AI assistant in retrieving and generating responses using RAG and web search. Evaluate how well the system meets the defined requirements and handles various user queries.
- **Code Quality** - Evaluate the structure, readability, maintainability, and efficiency of the codebase. Consider adherence to best practices, modularity, and proper documentation.
- **Non-Functional Requirements** – Additional features that improve user experience or provide additional functionalities are good, but not critical.