4. System Design and Modelling

This section outlines the overall design and modelling approach for the system that generates personalized UPSC mock interview questions using Generative AI and Large Language Models (LLMs). The system consists of the following component:

**Machine Learning Pipeline:**

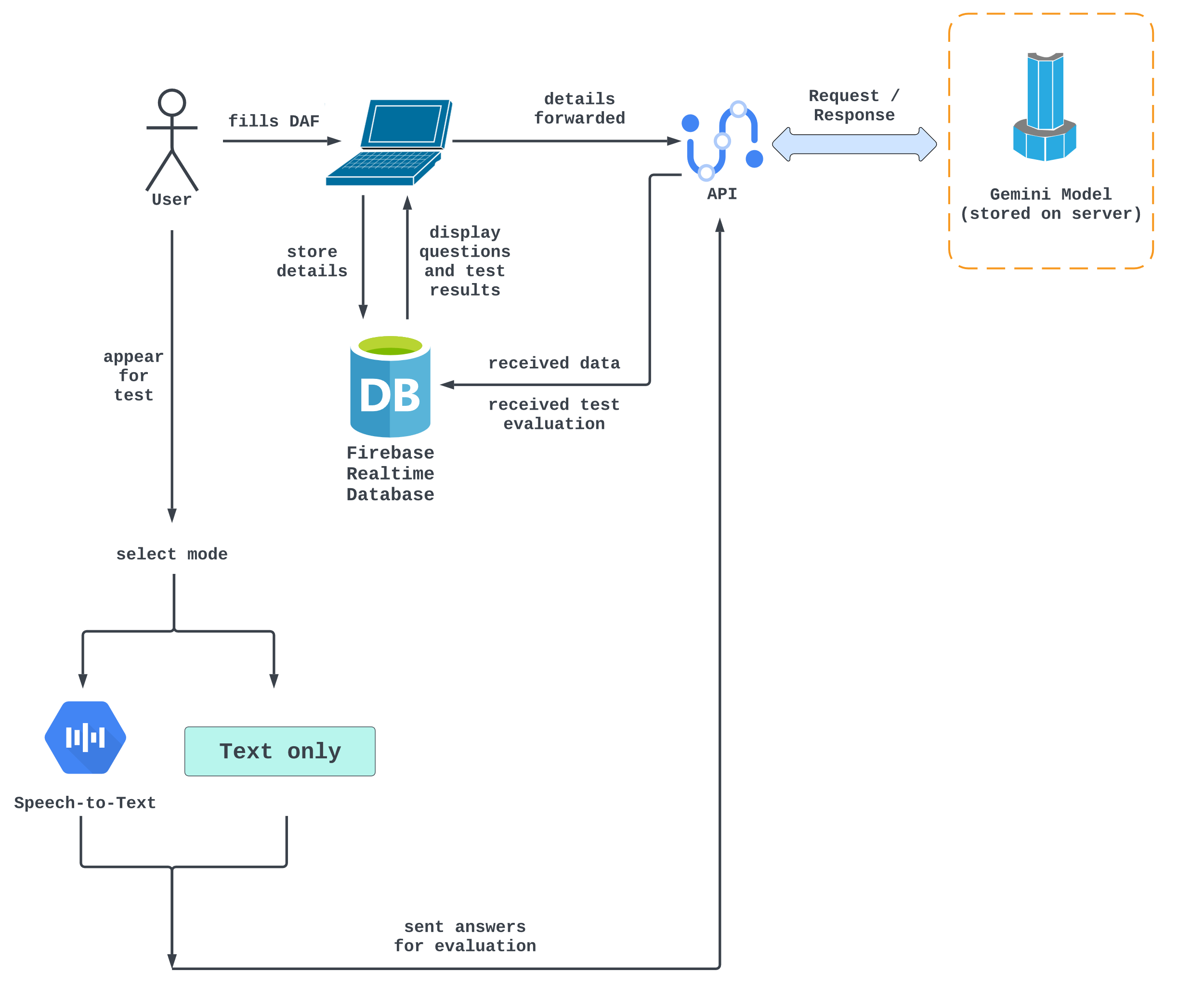
This pipeline encompasses the processes related to data preparation, model training, and question generation.

**Data Preprocessing:** The pipeline will begin by ingesting the UPSC interview question dataset and candidate DAF data. This data will undergo preprocessing steps such as cleaning, normalization, and tokenization to ensure consistency and facilitate efficient processing by the LLM.

**Model Training:** A pre-trained Generative AI model, potentially Google Gemini, will be employed for question generation. This model will be further fine-tuned on the preprocessed UPSC interview question dataset. The fine-tuning process allows the model to learn the patterns and styles of existing UPSC interview questions, enhancing its ability to generate relevant and realistic questions.

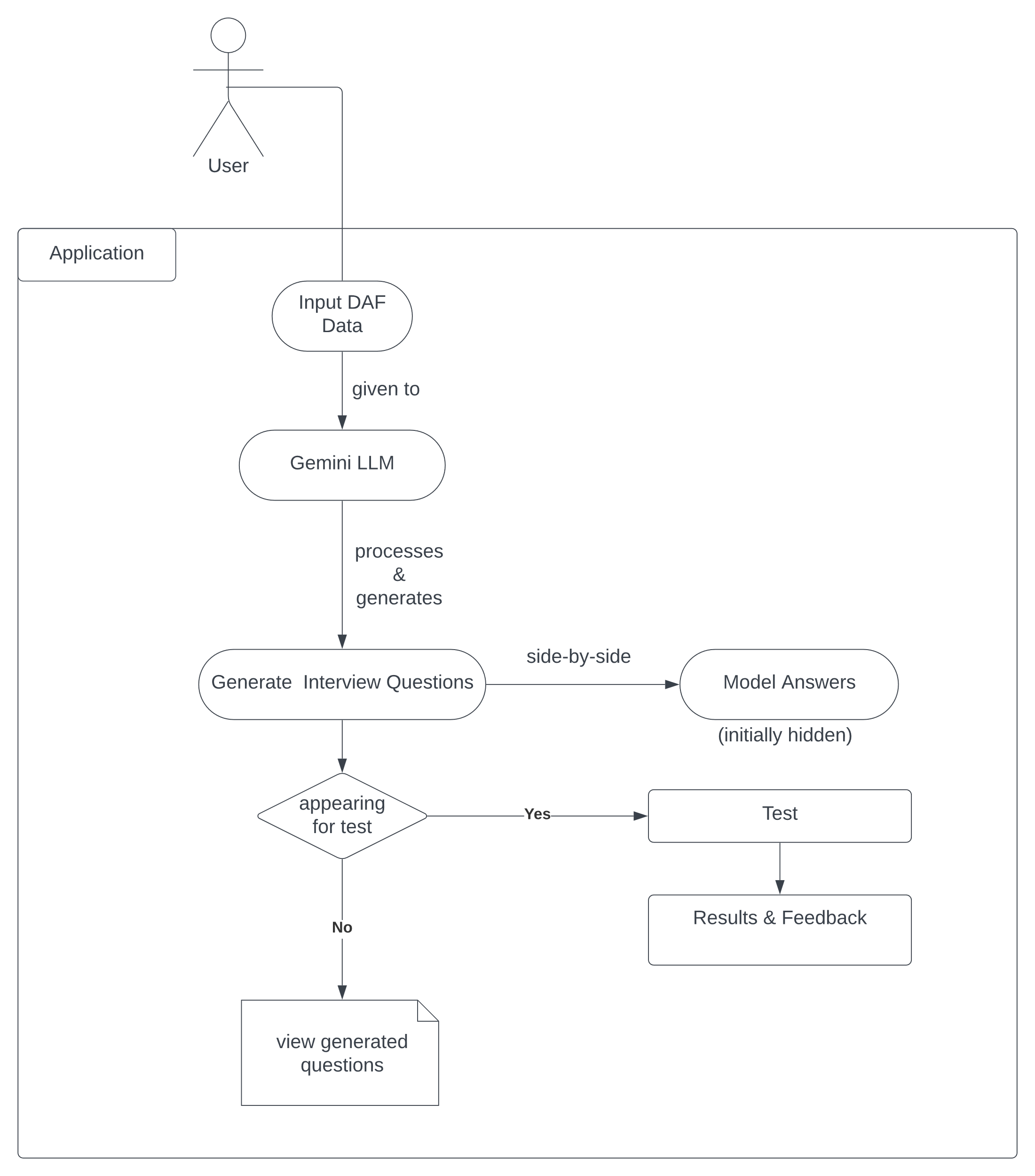
**Question Generation:** During the question generation stage, the preprocessed candidate data from the DAF will be fed into the fine-tuned LLM. The LLM, leveraging its knowledge of UPSC interview questions and the candidate's profile, will generate a set of personalized mock interview questions tailored to the candidate's strengths, weaknesses, and areas of experience.

**4.1 System Architecture**

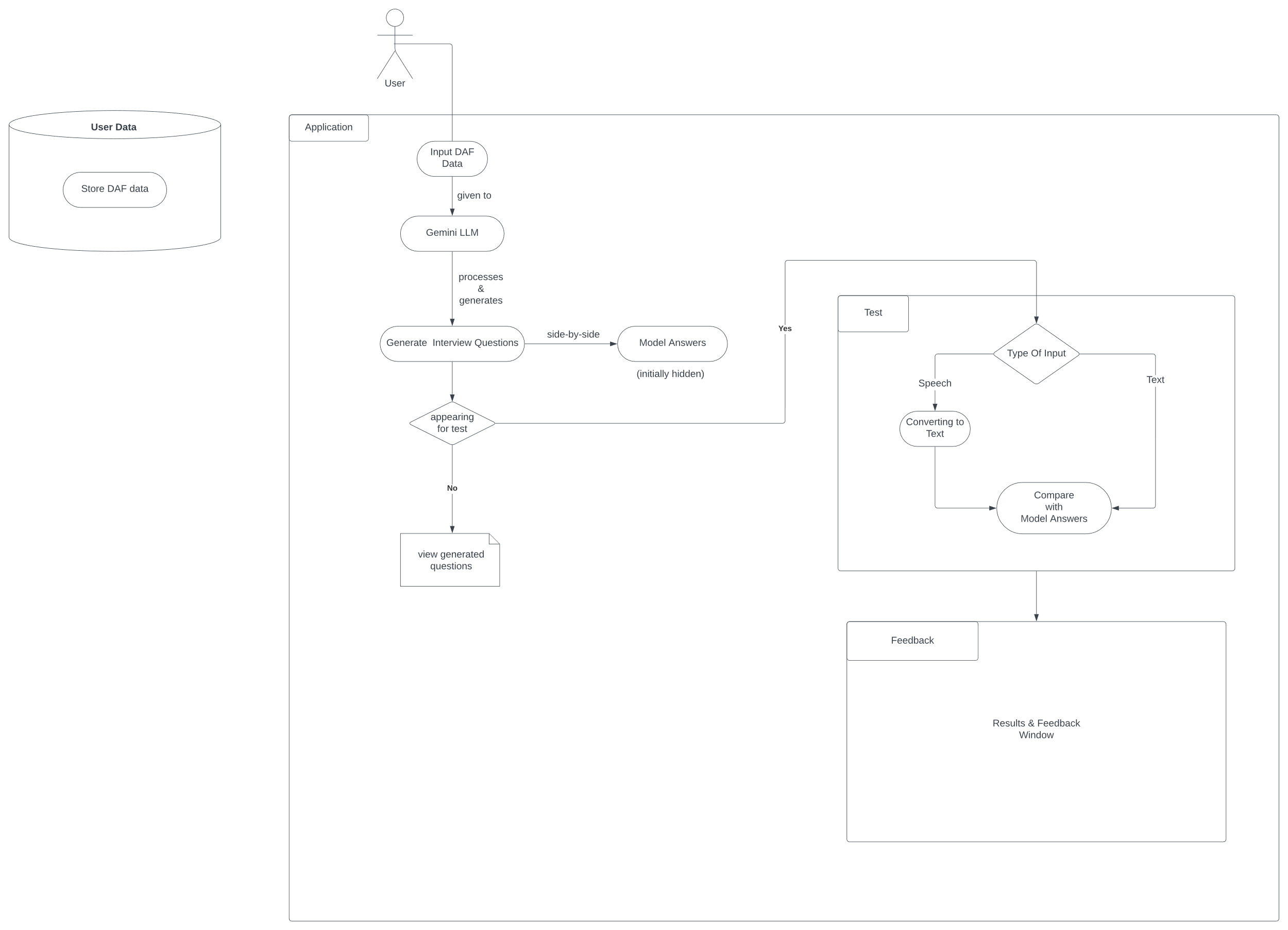


**4.1 Data Flow Diagrams**

**DFD Level 0 :**



**DFD Level 1 :**



**DFD Level 2 :**

