Report: Solution Overview

The solution addresses a challenge that focuses on using Large Language Models (LLMs) to interpret tabular data and provide clear and meaningful conclusions. The objective is to develop an LLM-based solution that can answer questions about events based on the provided tabular data.

Chosen Language Model (LLM):

The chosen Language Model (LLM) for this solution is the Gemini model version 1.0-pro provided by the Google GenerativeAI API. The Gemini model is a powerful LLM that can understand and generate human-like text responses based on given prompts. It leverages advanced natural language processing techniques to interpret and analyze the provided data and generate meaningful answers to the questions.

Architecture:

The solution follows a structured architecture to achieve the challenge objectives. It consists of the following components:

1. Data Ingestion: The solution ingests tabular data from a CSV file that is provided during the Submission phase. The CSV file contains the necessary data, including the questions to be answered and, in some cases, the expected answers.

2. Data Pre-processing: The chardet library is used to detect the encoding of the CSV file. This ensures that the data is read correctly, regardless of the encoding used. The detected encoding is then used to read the CSV file using the pandas library.

3. LLM-based Question Answering: For each question in the data file, the solution utilizes the Gemini LLM to compute a natural language answer based on the information provided with the question. The LLM analyzes the tabular data, interprets the context, and generates a response that provides a clear and meaningful conclusion.

4. Prompt Generation & Response Generation: The solution iterates over each row in the DataFrame obtained from the CSV file. It creates a prompt using the data from the row, including table page title, table section title, table array, and question. The Gemini LLM, specifically the GenerativeModel, is utilized to generate a response based on the prompt. The model analyzes the tabular data, interprets the context, and generates

a response that provides a clear and meaningful conclusion.

5. Answer Recording: The solution records the generated answers to the `solutions.csv` file. Each question corresponds to a row in the file, and the generated answer is stored in the appropriate column. This ensures that the answers are accurately associated with the respective questions.

Conclusion:

The provided solution demonstrates the utilization of the Gemini LLM to interpret tabular data and generate meaningful answers to questions about events. By leveraging advanced natural language processing techniques, the solution improves workplace efficiency by automating the analysis of large amounts of data and providing clear conclusions. The structured architecture ensures that the solution accurately processes the provided data and records the generated answers for evaluation.