Causal Analysis and Interactive Reasoning over Conversational Data

Problem Statement

Large-scale conversational systems generate extensive collections of multi-turn dialogue transcripts between agents and customers. Some of these conversations are associated with **explicitly labeled outcome events**, such as escalations, complaints, refunds, or other operationally significant events. These events are costly and important to analyze, but the **causal factors within the conversations** that lead to these outcomes are not directly observable.

Existing systems typically record that an event occurred for a given conversation, but do **not identify which specific dialogue turns or spans** contributed to the event, how conversational patterns evolve prior to the event, or which **recurring conversational structures** are systematically associated with similar outcomes across many conversations.

The challenge is to design a system that can operate over a **corpus of conversational transcripts** (with speaker labels and turn ordering, and optionally timestamps) and produce **causally grounded explanations** that link conversational behavior to observed outcome events. The system must work at scale, handle noisy conversational data, and produce outputs that are **interpretable and traceable back to concrete evidence** in the transcripts.

Beyond single explanations, users must also be able to **interactively explore these causal insights**. After receiving an initial explanation, users should be able to ask follow-up questions that depend on prior queries and system responses. The system must therefore **maintain contextual consistency** across multiple turns of interaction.

The overall objective is to move from **simple event detection** to **causal analysis and interactive reasoning over conversational data**.

Task 1: Query-Driven Causal Explanation with Evidence

Objective

Design a system that accepts a **single natural-language analytical query** related to a predefined outcome event and returns a **causal explanation grounded in conversational evidence**.

Task Definition

Given a corpus of conversational transcripts with **structured turns and speaker roles**, and a set of conversations labeled with outcome events, the system must answer user queries that ask **why a particular outcome event occurs**. The system must analyze the relevant conversations, identify **dialogue-level factors and patterns** associated with the event, extract **specific dialogue spans** that serve as supporting evidence, and produce an explanation that **explicitly links these factors to the outcome**.

Constraints

- The explanation must be **evidence-based** and reference identifiable portions of the data.
- The system must emphasize **causal reasoning** rather than simple correlation.
- The output must be **structured and interpretable**.

Task 2: Multi-Turn Context-Aware Query Handling

Objective

Extend the Task 1 system to support **multi-turn analytical interaction** while preserving contextual consistency.

Task Definition

Given an initial query and response, followed by one or more follow-up queries that depend on prior system outputs, the system must **retain context from earlier interactions**, correctly interpret follow-up questions, generate responses that **build upon previous analyses**, and maintain **consistency in causal reasoning and evidence usage**.

Constraints

- Context handling must be **explicit and deterministic**.
- The system must demonstrate **awareness of previously discussed events, conversational factors, and evidence**.
- All responses must be grounded in the **same underlying conversational data and causal framework**.

Judging Criteria

The model shall be evaluated on the following metrics:

- **ID Recall (Evidence Accuracy):** Measures whether retrieved Call IDs match ground-truth evidence.
- **Faithfulness (Hallucination Control):** Verifies that responses are strictly derived from retrieved context.

- **Relevancy (Conversational Coherence):** Evaluates whether the system directly addresses user intent, especially in multi-turn queries.

Teams should generate a **diverse set of queries** covering multiple domains and sufficient complexity to test follow-up reasoning. These queries must be included in the submission in `.csv` format with the following columns:

`Query_Id, Query, Query_Category, System_Output, Remarks`.

Deliverables

System Implementation

- A **complete end-to-end implementation** covering both Task 1 and Task 2.
- All submissions must be provided in a **`.zip`** file, including:
  - Source code
  - Codebase for Task 1 and Task 2 with trained model(s)
  - `requirements.txt` listing all external libraries and versions
- A **technical report** explaining the approach and evaluation results along with the query dataset.
- A **well-structured and reproducible** codebase.

General Guidance

README File: The submission must include a **comprehensive `README.md`** file providing clear, step-by-step instructions to set up the environment and run the system, including data preprocessing scripts, model training, and execution of the final application.