

# Causal Conversational AI System Report

## Problem Statement

Large conversational systems generate multi-turn dialogues between agents and customers. Some conversations result in operational events such as escalations or complaints. Traditional systems detect these outcomes but fail to explain which dialogue segments caused them. The objective is to build an AI system that performs causal reasoning over transcripts, extracts supporting evidence, and allows interactive multi-turn analytical queries while maintaining context.

## Model Architecture

- 1 Transcript ingestion and structured turn parsing
- 2 Sentence embedding using transformer-based encoder
- 3 Event-conditioned logistic regression classifier
- 4 Turn-level causal scoring
- 5 Evidence extraction linked to transcript IDs
- 6 Structured explanation generator
- 7 Context memory for multi-turn reasoning

## Evaluation Metrics

- 1 ID Recall — measures correctness of retrieved transcript evidence
- 2 Faithfulness — ensures outputs derive strictly from transcripts
- 3 Relevancy — checks conversational coherence in follow-up queries
- 4 Deterministic context consistency — maintains reasoning continuity

## Results Summary

The system successfully identifies high-risk conversational turns contributing to escalation events. Evidence is traceable to transcript IDs, enabling transparent causal explanations. Multi-turn interaction preserves prior context and supports follow-up analytical queries. The model demonstrates scalable, interpretable reasoning suitable for conversational analytics workflows.