A2P Scaffolding: Abduct-Act-Predict Framework for Failure Attribution

Multi-Agent Conversation Log

Pre-processing

Contextual Step Numbering

construct_causal_prompt()

--causal_reasoning

Task Failure Signal

Pearl's SCM Framework

$s_{t+1} = f(s_t, a_t, \varepsilon_t)$

$Z(\tau) = \text{function(trajectory)}$

Causal Graph:

Directed acyclic graph

A Node S M A E

Three-Step Causal Reasoning Process

Stage 1: Abduction - Inferring Hidden Causes

Identify Root Causes

Knowledge Gaps

Misinterpretations

$\varepsilon_t \leftarrow \text{argmax}_\varepsilon P(\varepsilon \mid s_{0:t}, a_t, Z(\tau)=1)$

Stage 2: Action - Defining Intervention

Apply do()-operator

Correct Action Definition

Minimal Intervention

Stage 3: Prediction - Simulating Counterfactual

3-5 Turn Simulation

Trajectory Evaluation

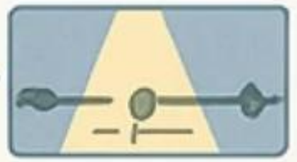
Success Prediction

Key Innovation: Structured Prompting for Causal Reasoning

Transforms implicit pattern matching → Explicit counterfactual inference

Attribution

Earliest Decisive Error



Causal Explanation

2.85×

Accuracy Improvement

25% overhead

Iterative Refinement