Technical Appendix

A: Prompts Used in the CIRF

(1) FOL Reasoning Generation: For each sentence-target pair, we first prompt an LLM to generate a reasoning chain in the form of first-order logic (FOL) (*P1*).

> $P1$: Your task is to analyze the attitude of the [*given sentence*] towards the [*given target*] using first-order logic. Formulate a response and conclude with a statement indicating the attitude (Support, Opposed, Neutral).

(2) FOL Interpretation and Abstraction:

Then, we further prompt the LLM to analyze their internal structure and inference logic (*P2*) (logical structure) and to produce alternative but logically equivalent FOL expressions (*P3*) (inference pattern). This step enables us to identify the underlying reasoning strategy, independent of surface form, and prepares the basis for subsequent abstraction.

> $P2$: Given a [*FOL expression*], analyze its reasoning structure, describe its logical form, and explain how it infers the attitude expressed in the text toward the target.

> $P3$: Can you provide alternative first-order logic (FOL) formulations that represent the same reasoning process but in different structural forms?

Next, we analyze multiple logically similar but structurally different FOL formulations. Based on these variations, we ask the LLM to summarize the shared reasoning pattern into a generalized FOL template (*P4*). This abstract form captures the essence of the reasoning logic in a more reusable and domain-independent way.

> $P4$: You are an expert in formal logic and natural language inference. Your role is to generalize reasoning patterns in the form of first-order logic (FOL) across variations of inference processes used for classification tasks. Based on [*logical structure*] and [*inference pattern (alternative FOLs)*], can we derive a general FOL reasoning template for inference task?

(3) Schema Clustering via LLM-Guided Reasoning Classification: With these abstracted logic forms, we perform clustering not only based on surface similarity, but also on underlying logical structure. We prompt the LLM to group FOL templates into logical categories or "schemas" based on inference patterns (*P5*). If a new logic form does not fit any existing schema, the LLM creates a new one with a unique ID, name, and description.

To support schema generalization (*P6*), we generate abstract logical schemas that summarize inference patterns across clusters of logic forms (FOLs). For each cluster, we prompt the LLM to synthesize a representative logic chain that captures the shared reasoning structure, using both the cluster's description and its constituent FOLs.

When a cluster contains a large number of FOLs, we adopt a hierarchical abstraction strategy. FOLs are first grouped into sub-clusters based on fine-grained logical similarity. Each sub-cluster is summarized into an intermediate logic chain. And in final, these intermediate summaries are then merged to form the final schema-level logic chain.

> $P5$: You are a reasoning classification expert specialized in identifying logical patterns in First-Order Logic (FOL). When given a [*FOL expression*], classify it based solely on its [*logical structure*], [*inference pattern*], and [*generalized form*]. If it matches an existing abstract cluster, assign it accordingly. If not, create a new cluster with id, name, and description.

> $P6$: You are a logical reasoning summarizer and generalizer. Your job is to analyze a cluster of structurally similar first-order logic (FOL) formulas that represent individual reasoning processes for the same task. Please summarize the cluster of FOL based on their shared logical structure and inference pattern.