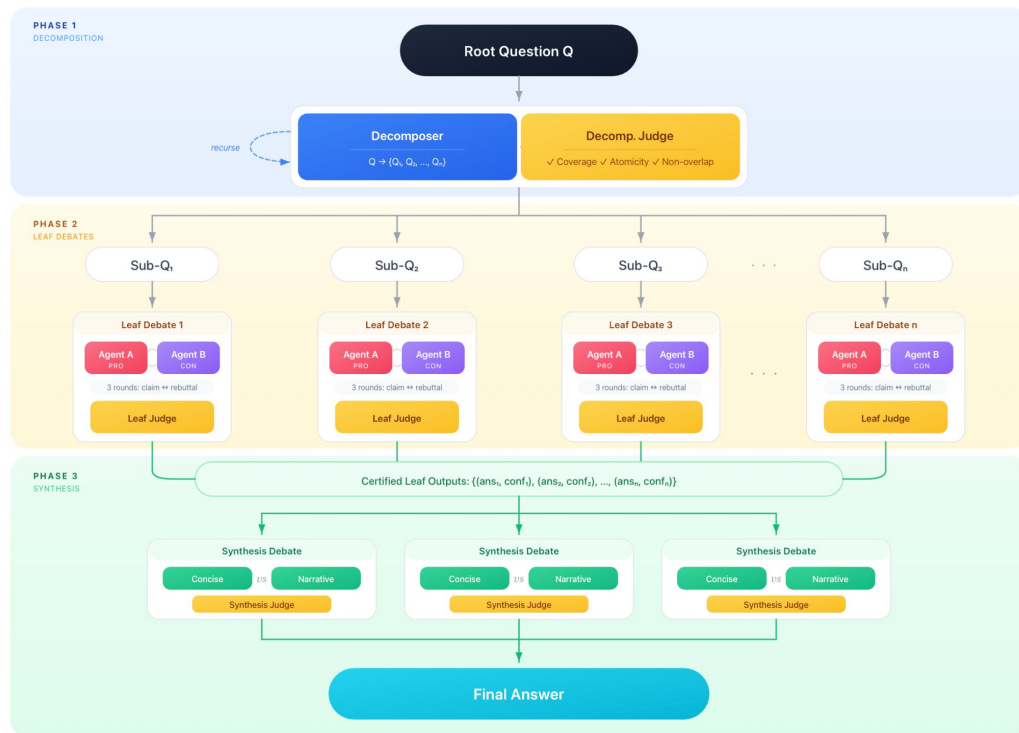# Coordinating LLMs via Debate Trees: Hierarchical Decomposition Improves Truthfulness

Xiang Fu and Kevin Gold
Faculty of Computing and Data Sciences, Boston University

BOSTON
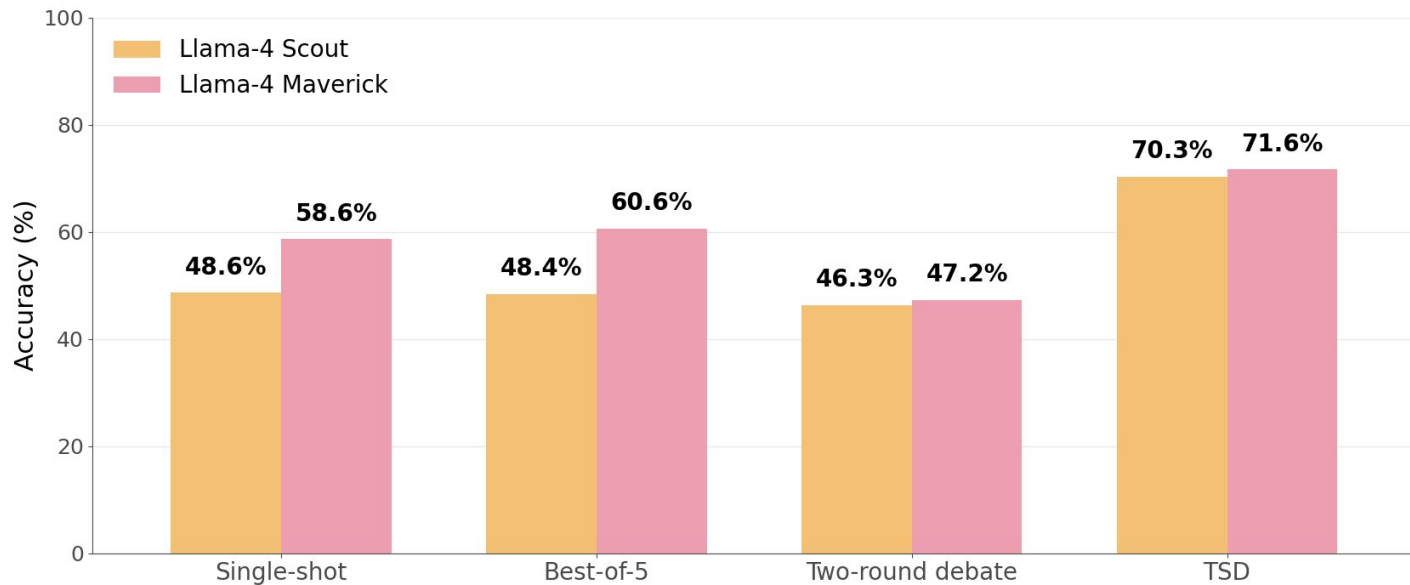UNIVERSITY

Tree-Structured Debate

Decompose: split Q into 2 to 4 non-overlapping sub-questions (judge checks coverage, atomicity, non-overlap).

Leaf debates: run parallel debates on atomic leaves; a leaf judge certifies each local answer.

Synthesize: integrate certified leaves bottom-up to produce the final answer.

Disagreements become local, auditable, and fast via parallel leaves.

TruthfulQA Results

Why Trees Help: Exposing Hidden Assumptions

Root question: Which language do all people in Europe speak?

Leaf Q1: Is there a single language spoken by everyone in Europe?
Answer: No.

Leaf Q2: Which languages are widely used across many European countries?
Answer: English is the most widespread cross-border lingua franca.

Leaf Q3: Within a given country, does everyone share the same language?
Answer: No; many countries are multilingual.

Takeaway. Decomposition separates universal claims, prevalence, and within-country variation. This prevents mixing "common" with "universal" when answering complex questions. It also keeps critique focused on the specific assumption that fails.