

Coordinating LLMs via Debate Trees: Hierarchical Decomposition Improves Truthfulness

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



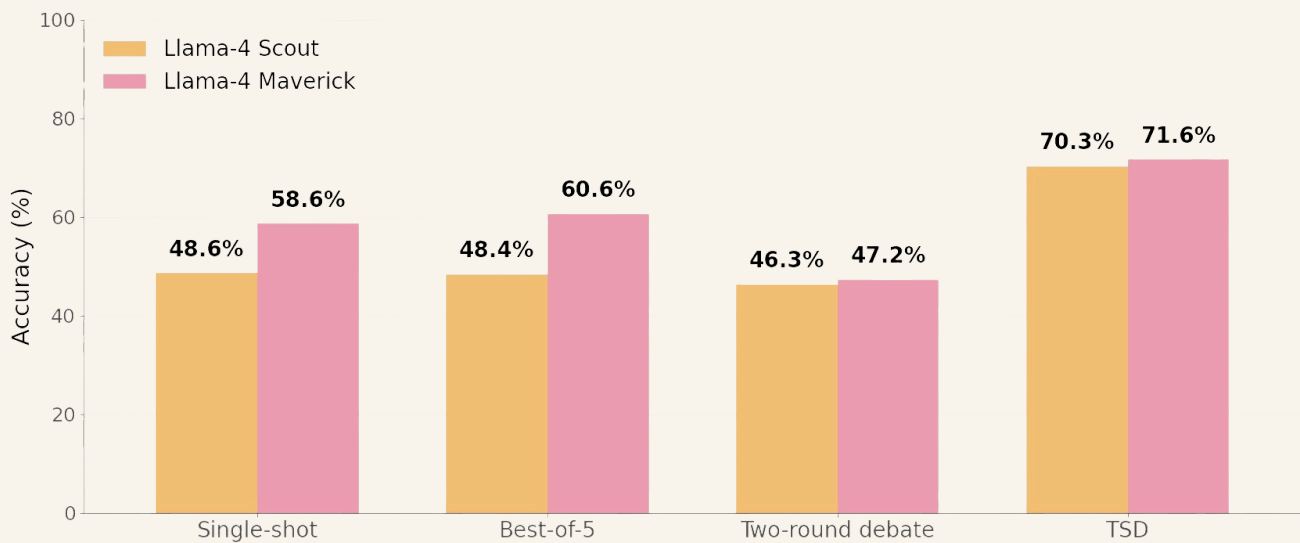
Why TSD?

Flat debate mixes many sub-issues in one thread, so critiques can miss key misconceptions.

Key Results (TruthfulQA)

70.3% meta-llama/llama-4-scout

71.6% meta-llama/llama-4-maverick



Takeaway

Localize disagreements to single branches

Parallel leaf debates keep latency practical

Bottom-up synthesis yields an auditable final answer