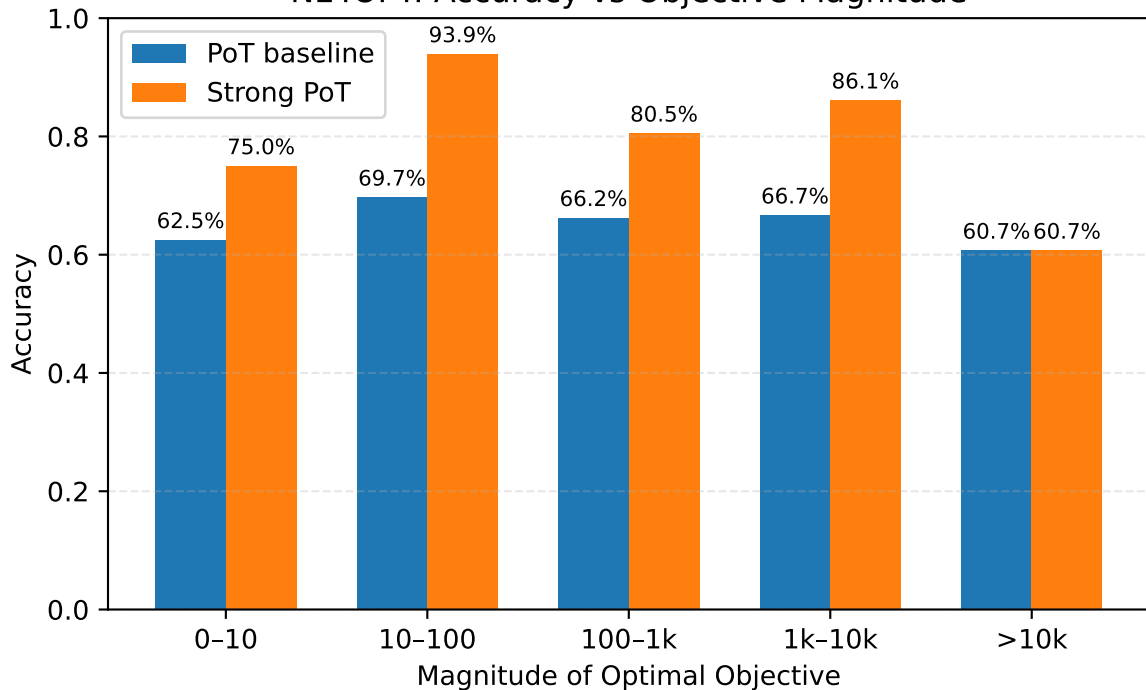
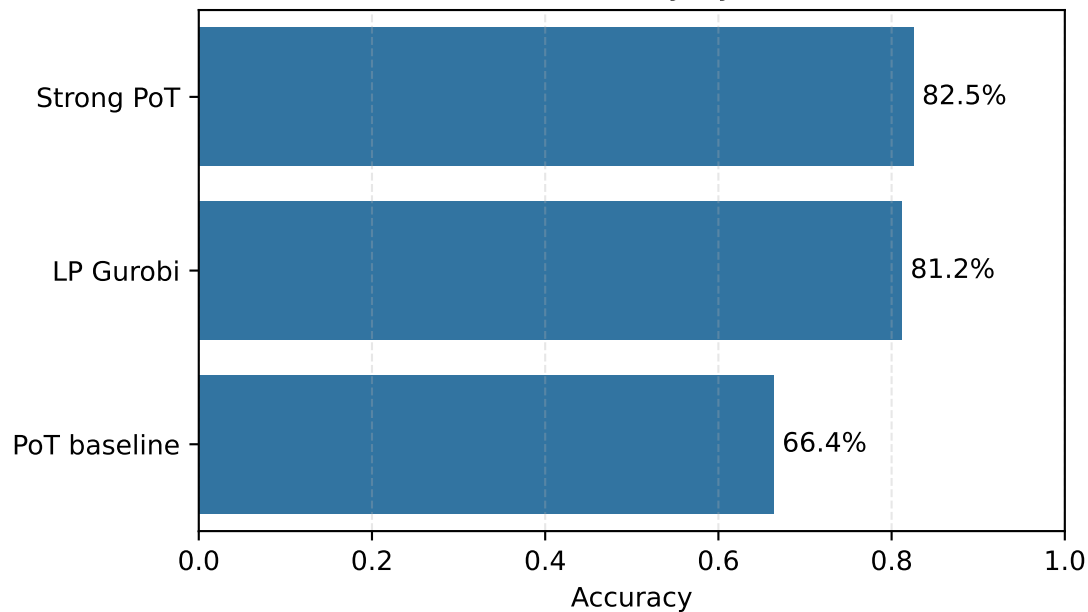


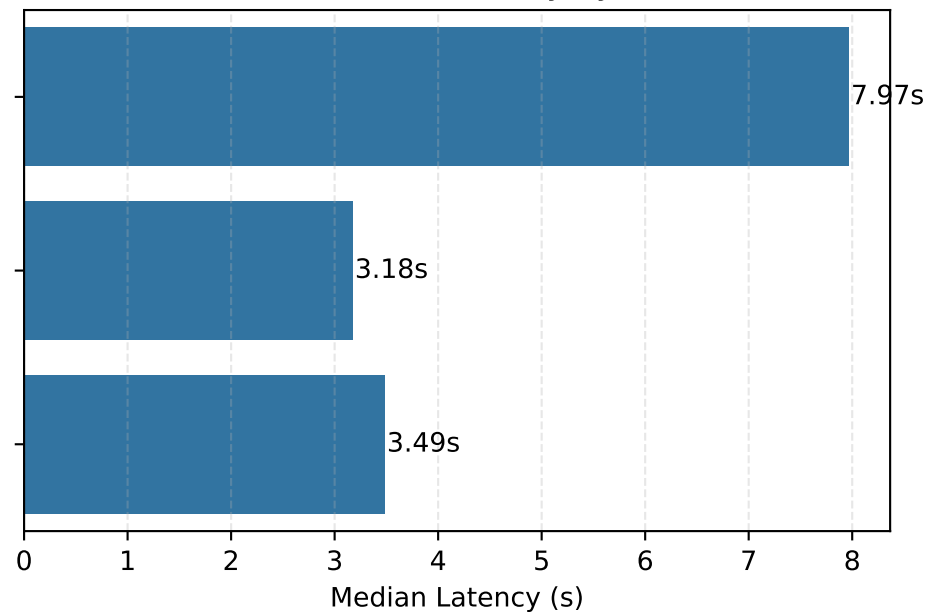
# NL4OPT: Accuracy vs Objective Magnitude



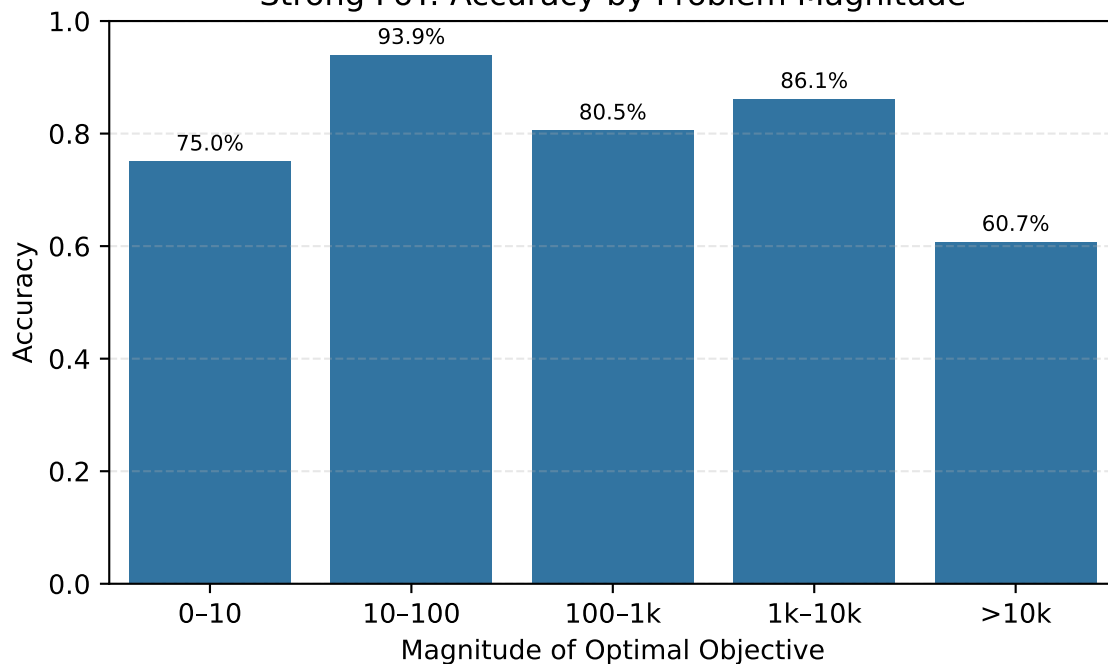
NL4OPT: Accuracy by Method



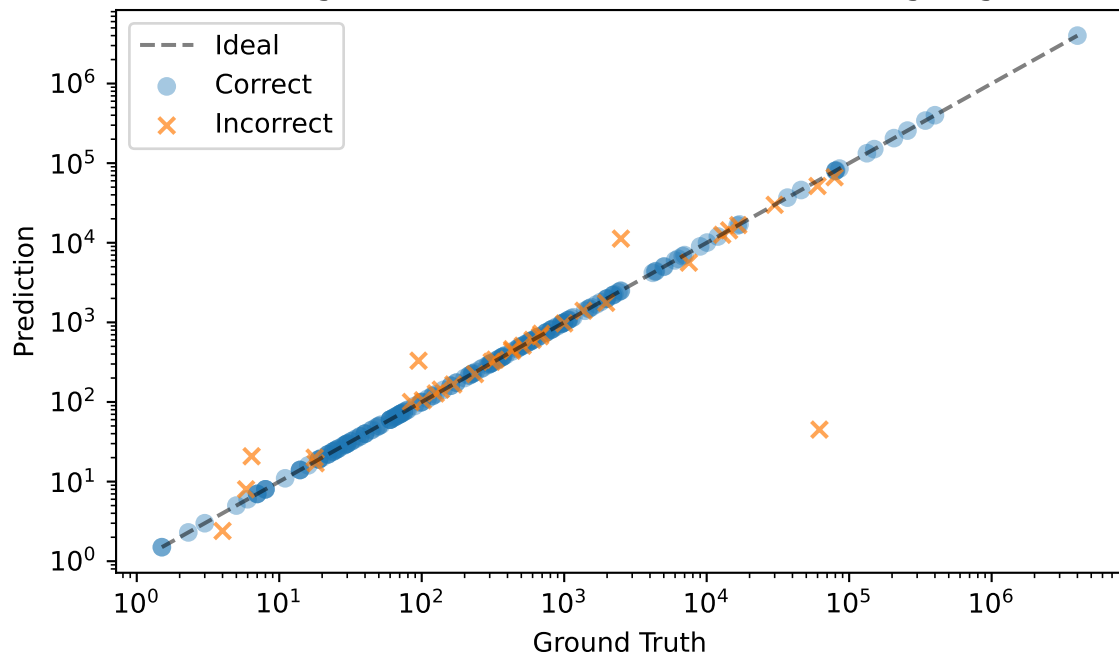
NL4OPT: Median Latency by Method



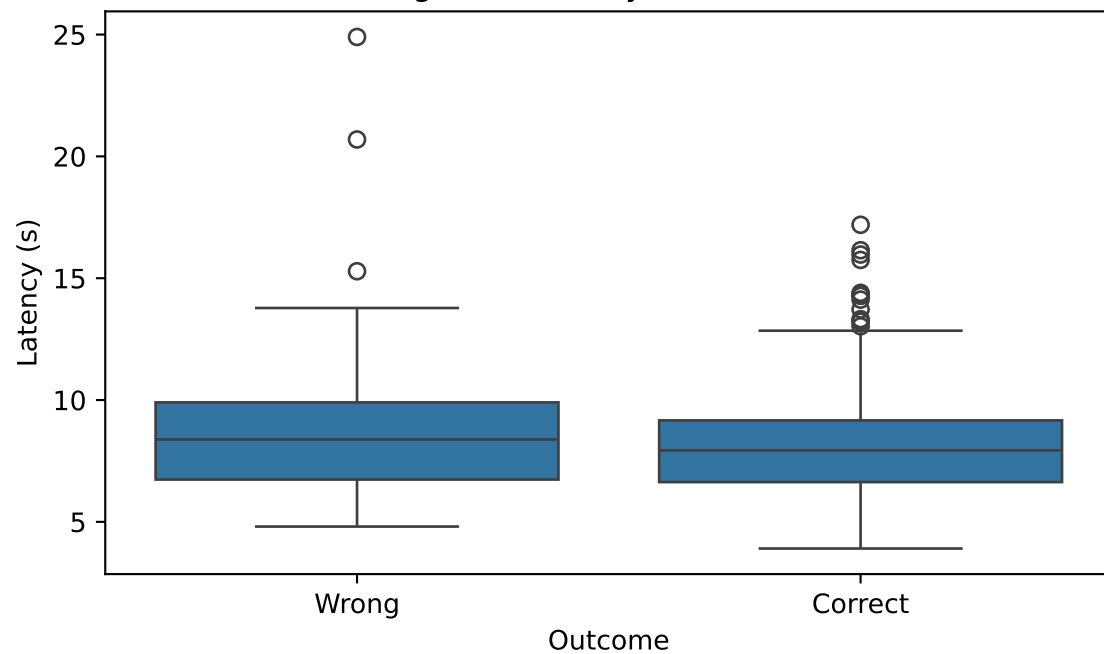
### Strong PoT: Accuracy by Problem Magnitude



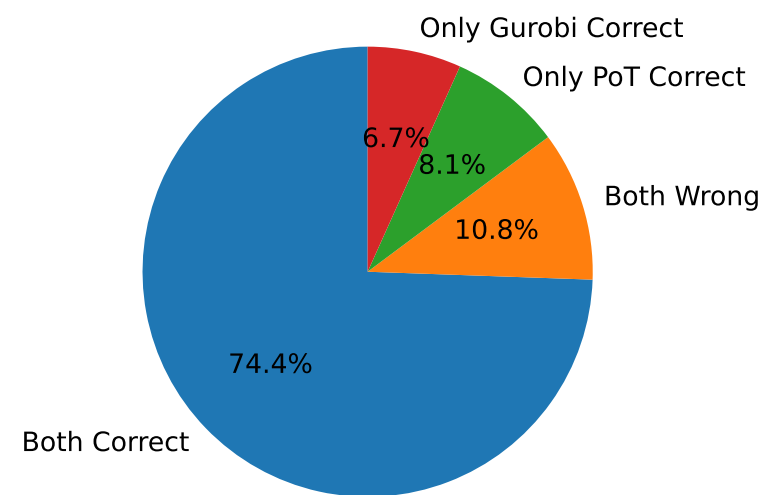
### Strong PoT: Prediction vs Ground Truth (Log-Log)



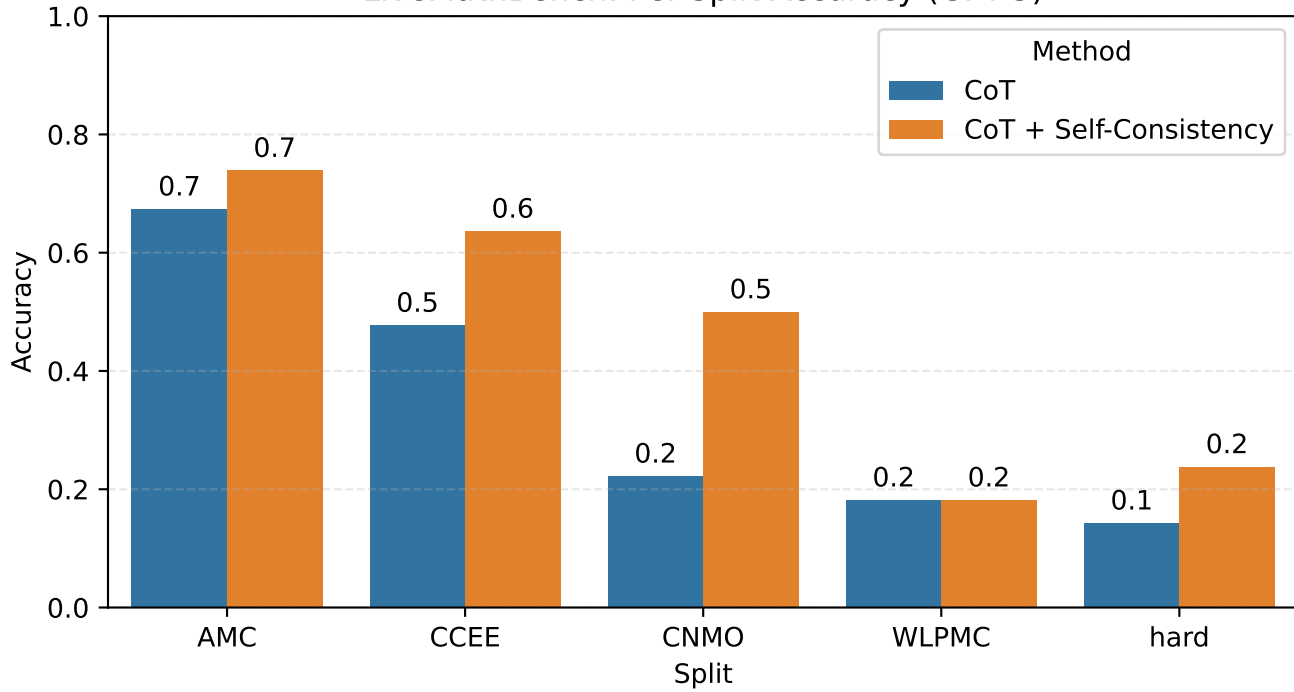
### Strong PoT: Latency vs Correctness



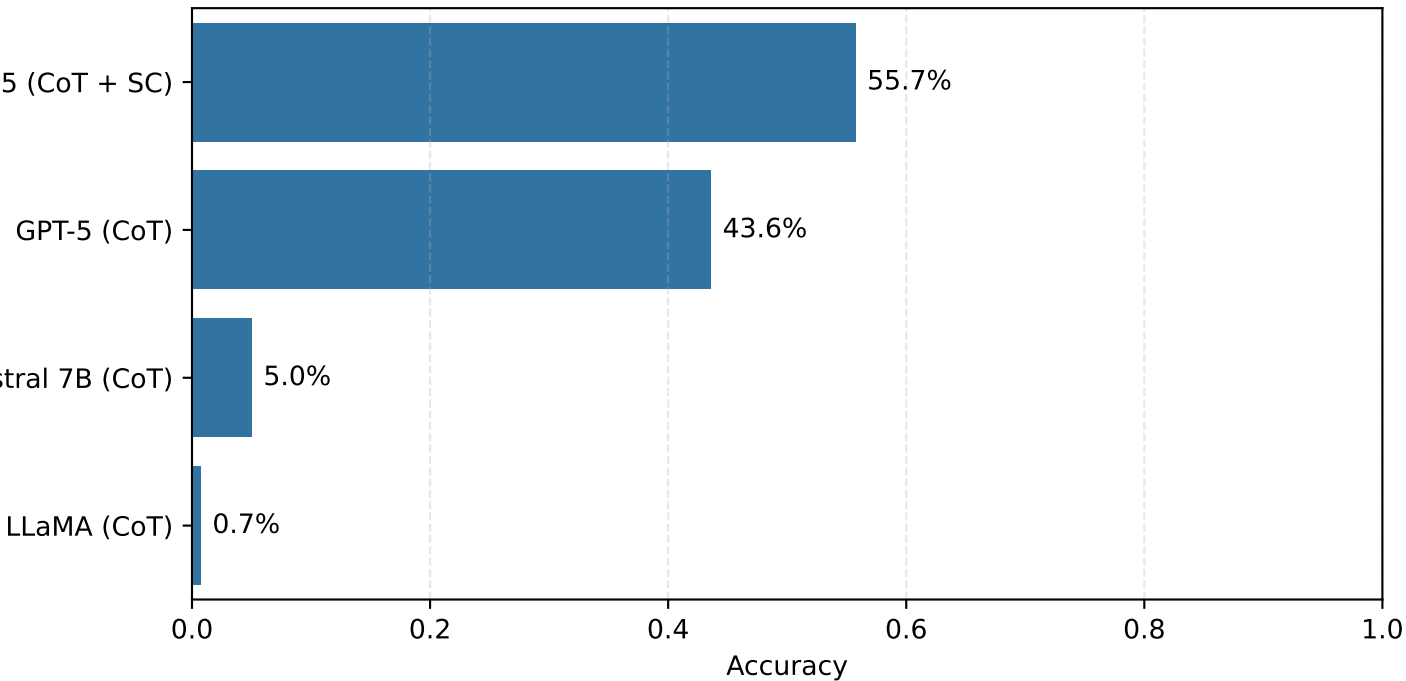
### Strong PoT vs Gurobi: Overlap of Success



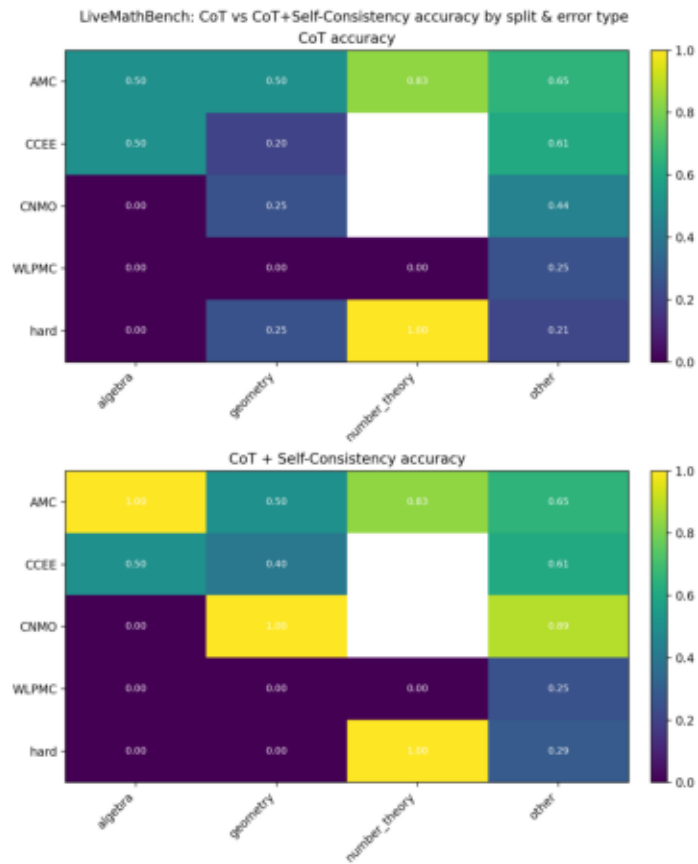
LiveMathBench: Per-Split Accuracy (GPT-5)



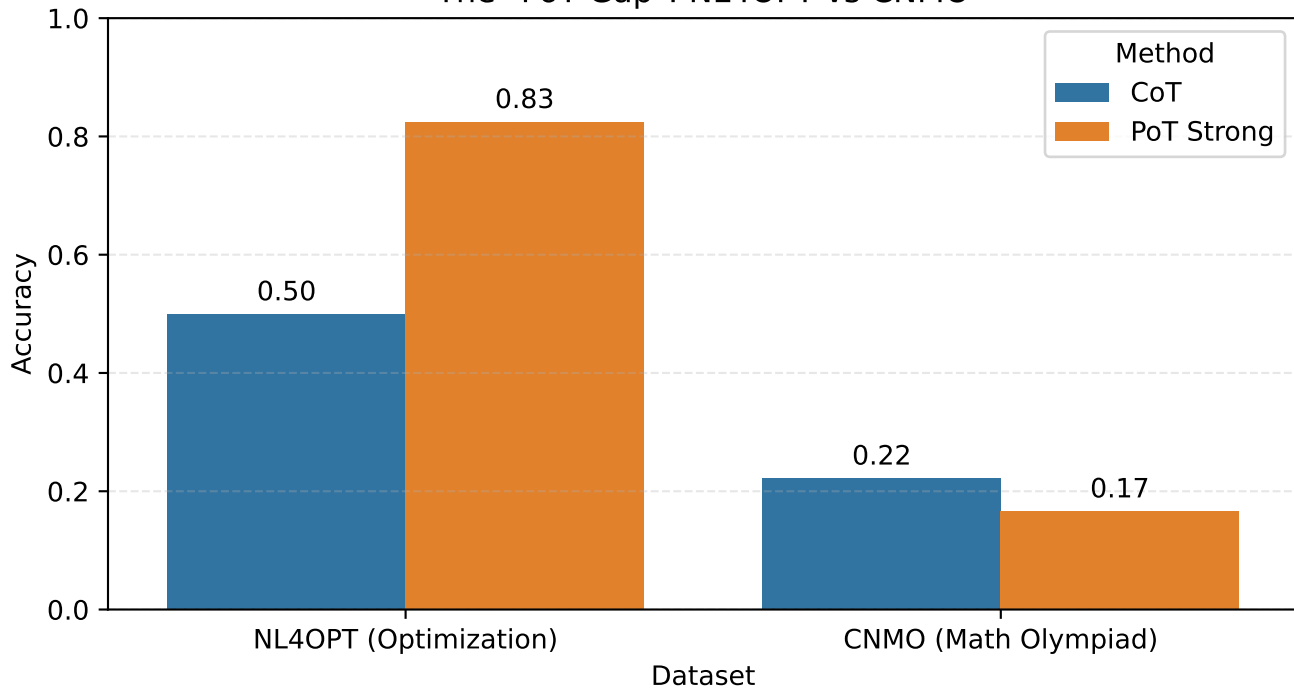
LiveMathBench: Overall Model Leaderboard



# LiveMathBench: CoT vs CoT+SC (Split × Topic)



The "PoT Gap": NL4OPT vs CNMO



Router Performance vs Task Mix

