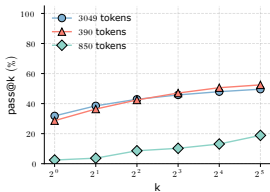
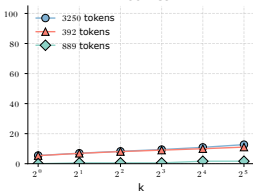


Thinking    NoThinking    Qwen-Instruct

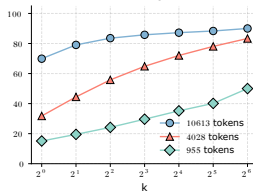
MiniF2F



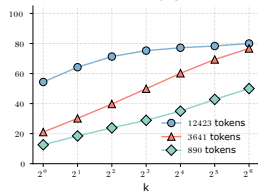
ProofNet



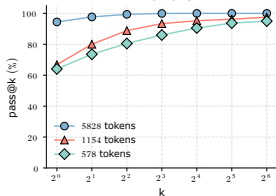
AIME 2024



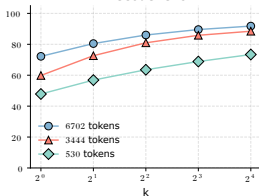
AIME 2025



AMC 2023



LiveCodeBench



OlympiadBench-Math

