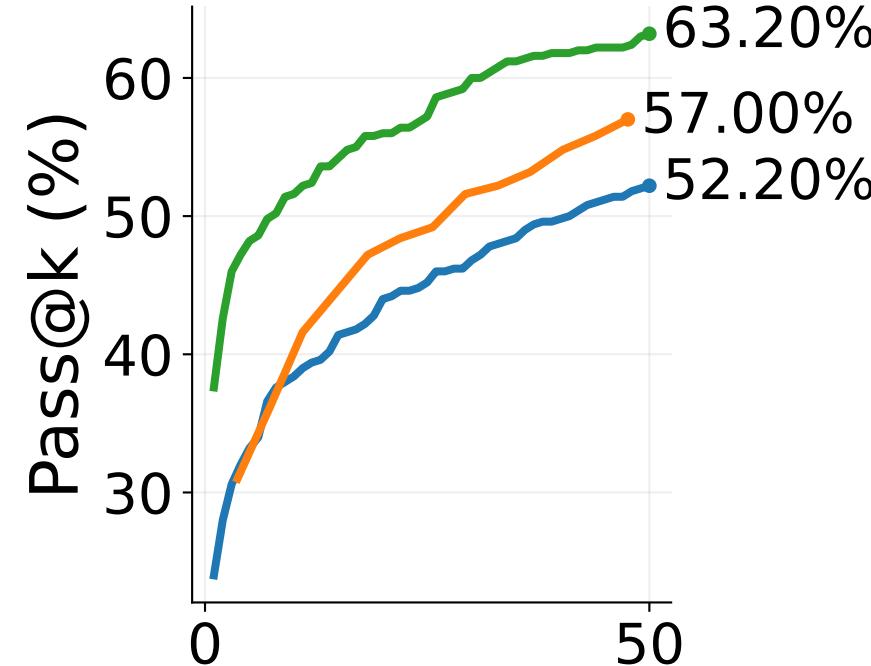
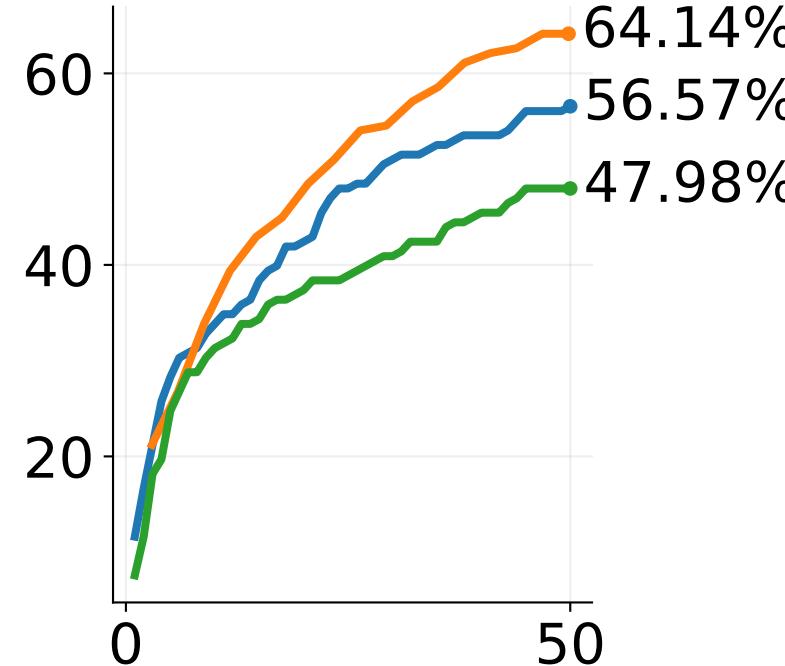


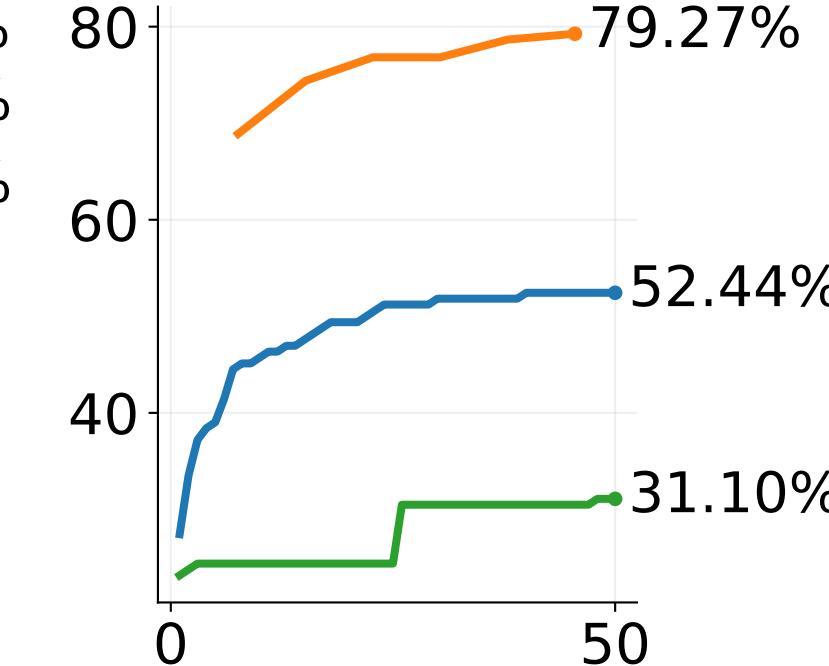
MATH



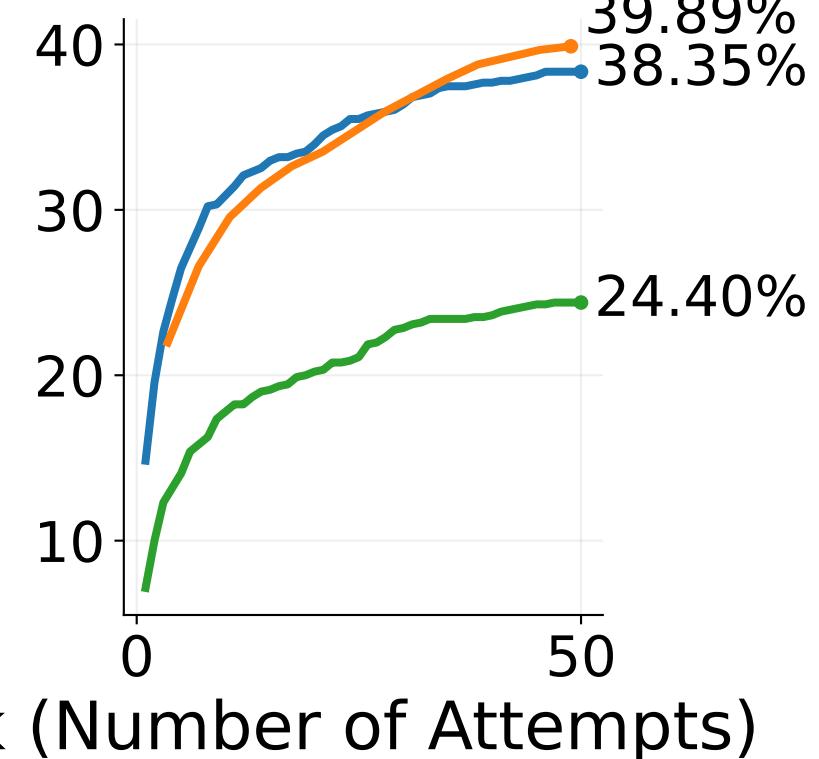
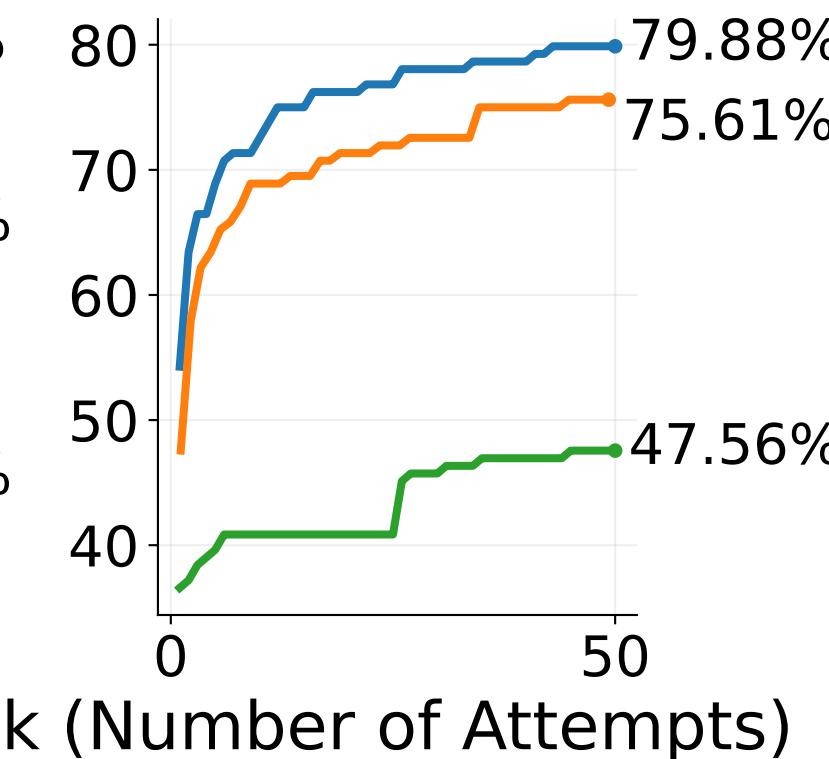
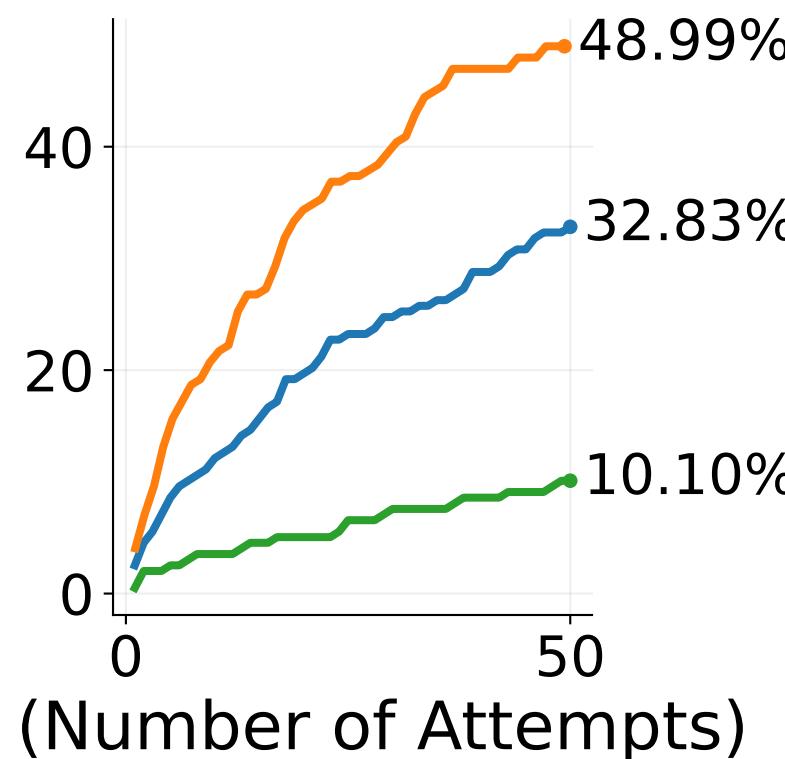
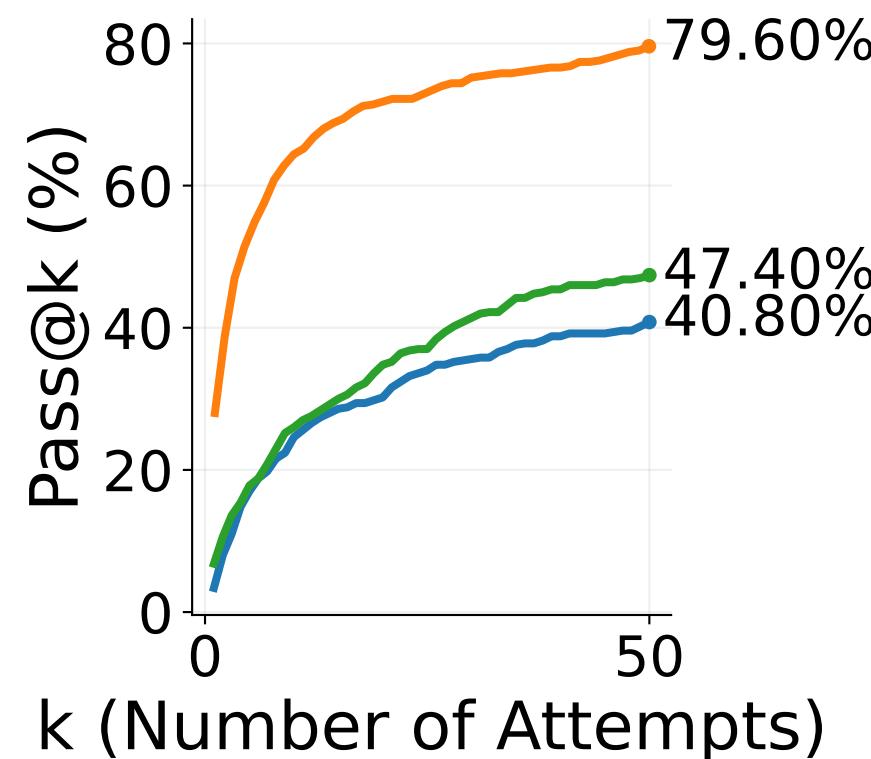
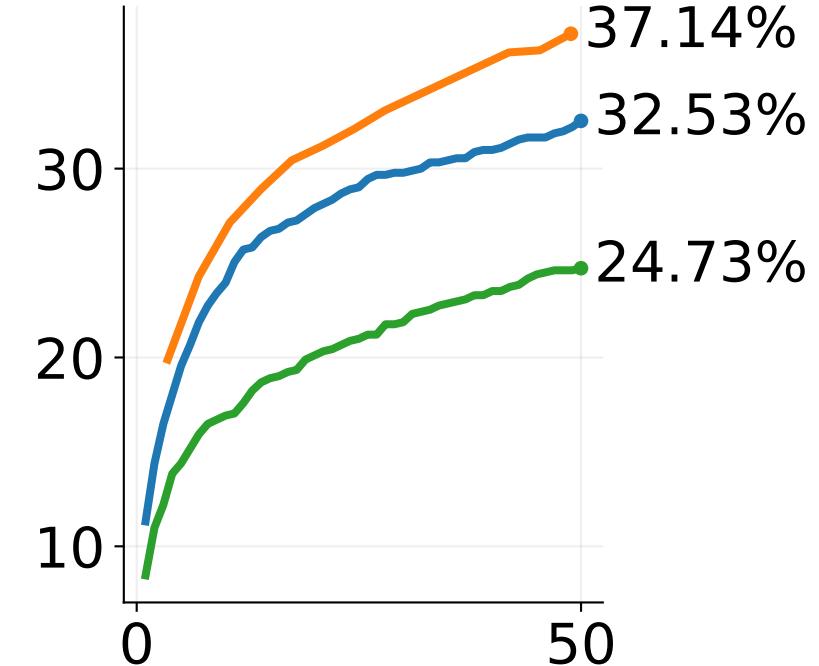
GPQA-Diamond



HumanEval



Olympiad Bench



— Traditional Repeated Sampling  
— GUIDED SAMPLING  
— Tree-of-Thought