

DeepSeek-R1

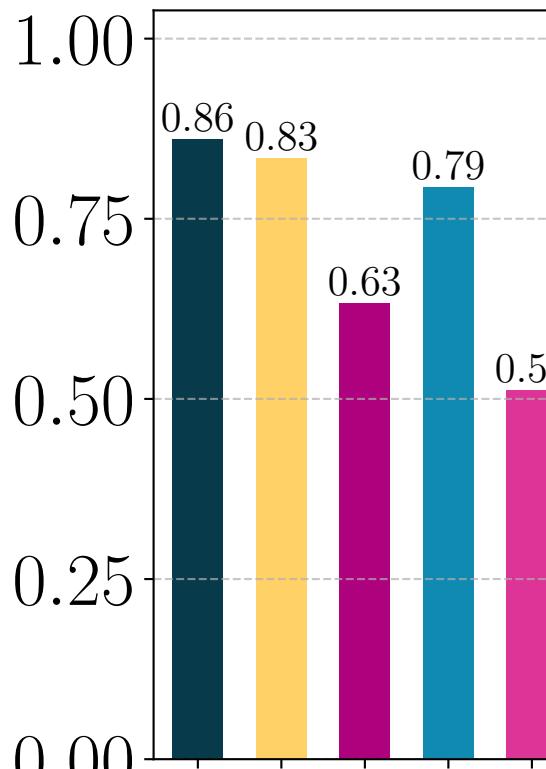
Grok 3 Mini Beta

Claude 3.7 Sonnet

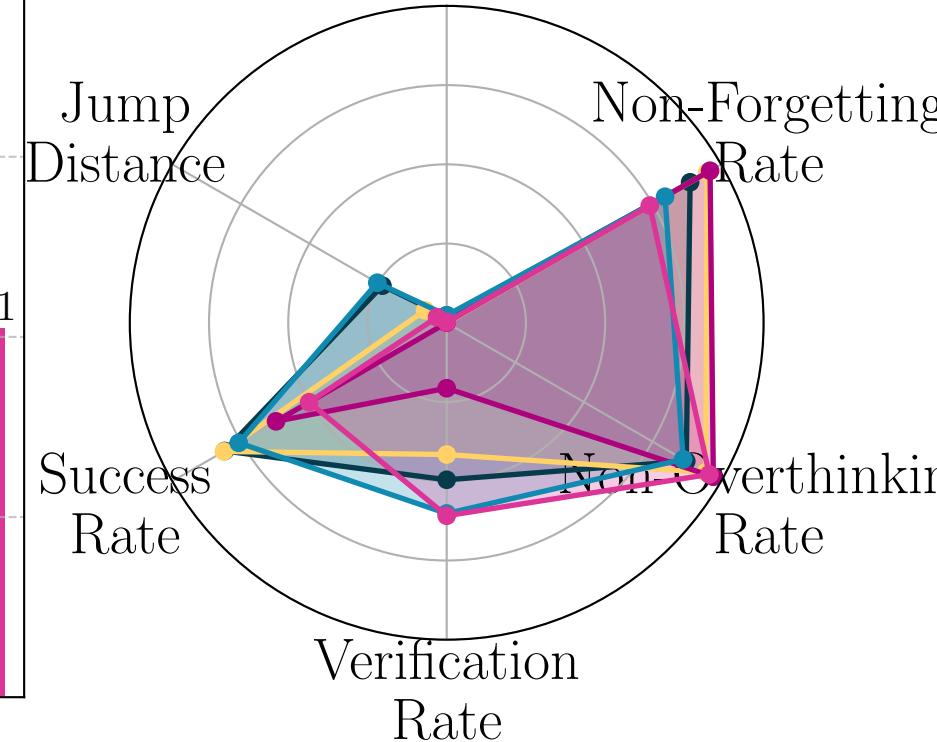
QwQ-32B

Phi-4-reasoning-plus

Pass@1

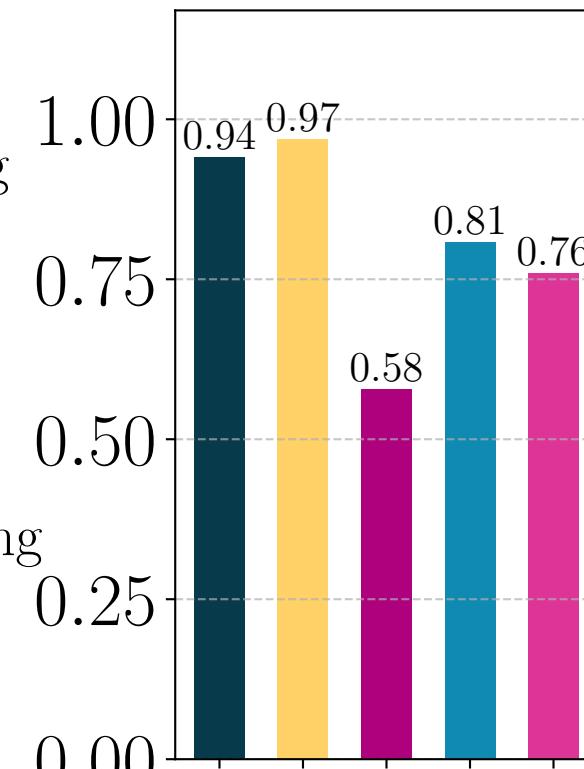


Solution Count

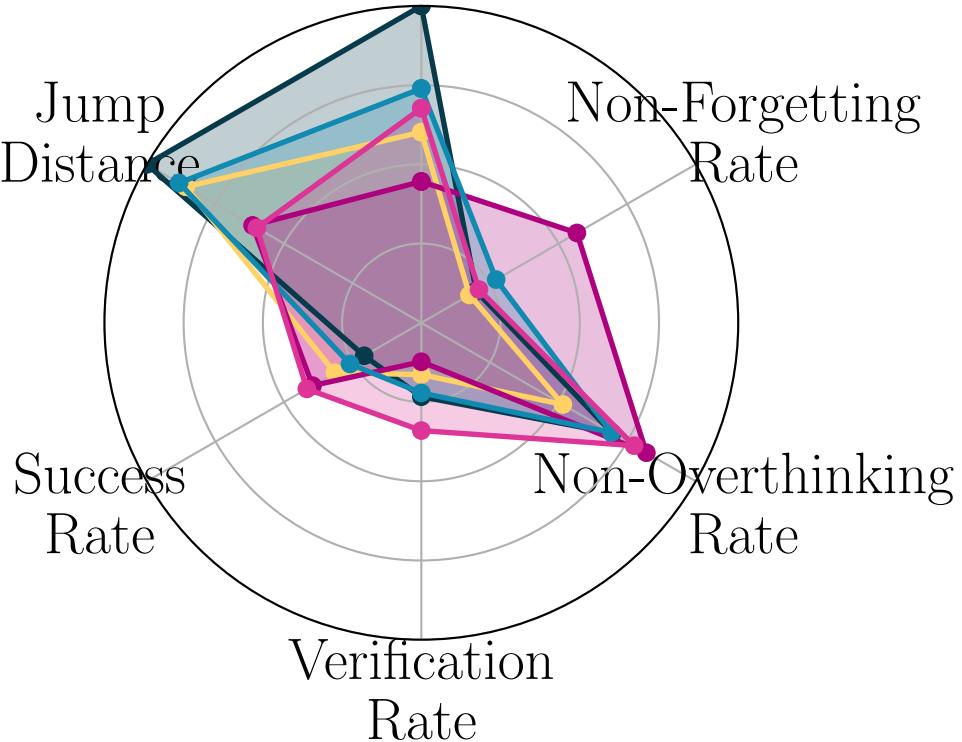


MATH500

Pass@1



Solution Count



Game of 24