

Gemini 2.0 Flash:

No Reasoning Examples 1

Reasoning Example

ng Examples 3 Reasoning Exam

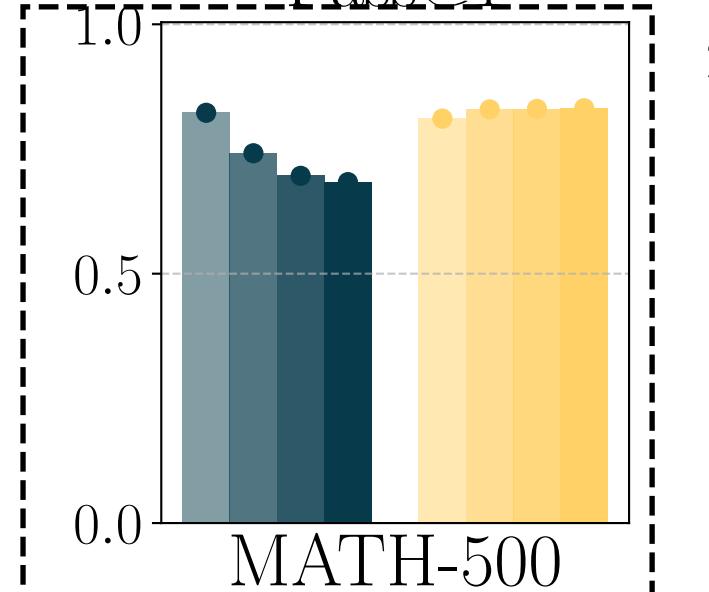
DeepSeek-V3:

No Reasoning Examples 1

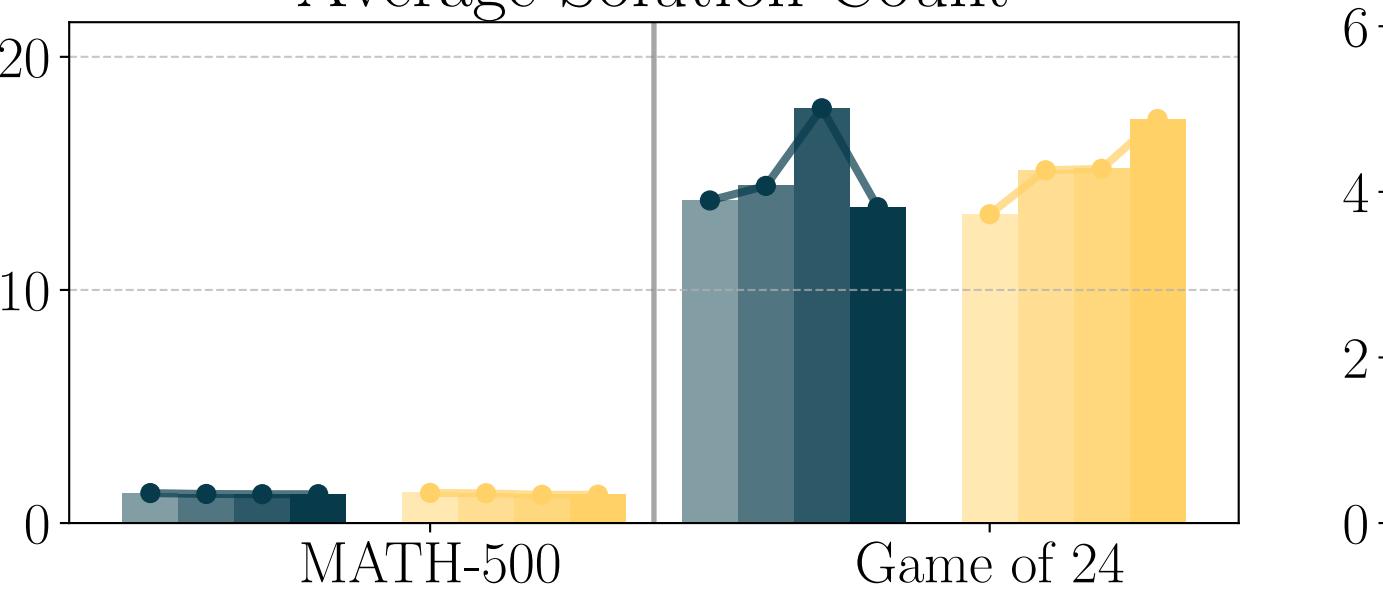
Reasoning Example 2 Reasoning

Reasoning Examples

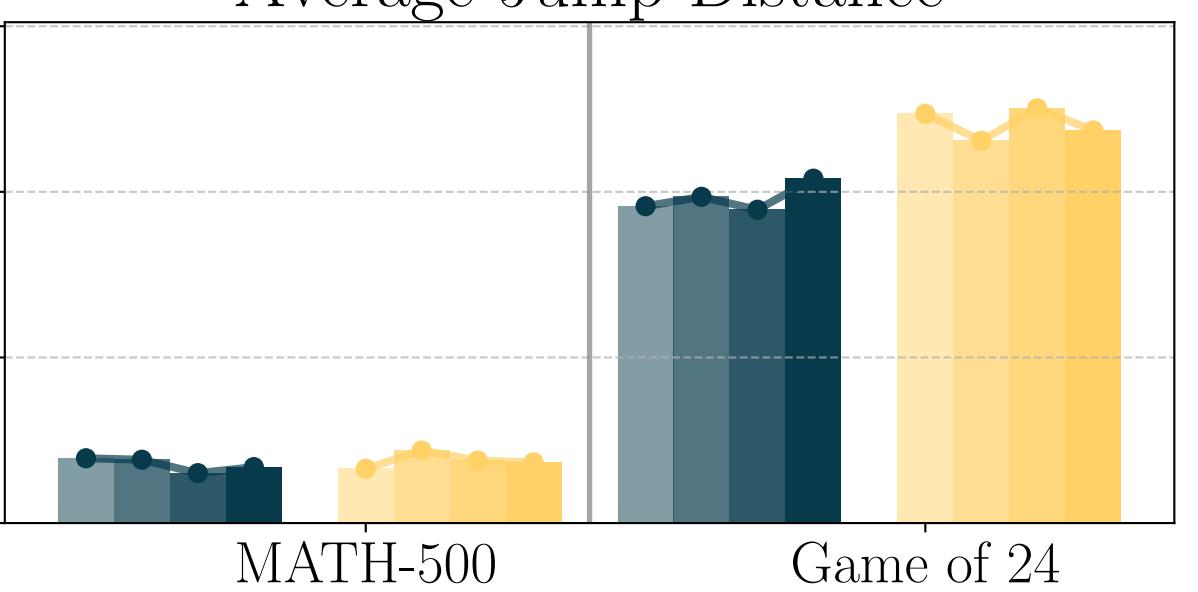
Pass@1



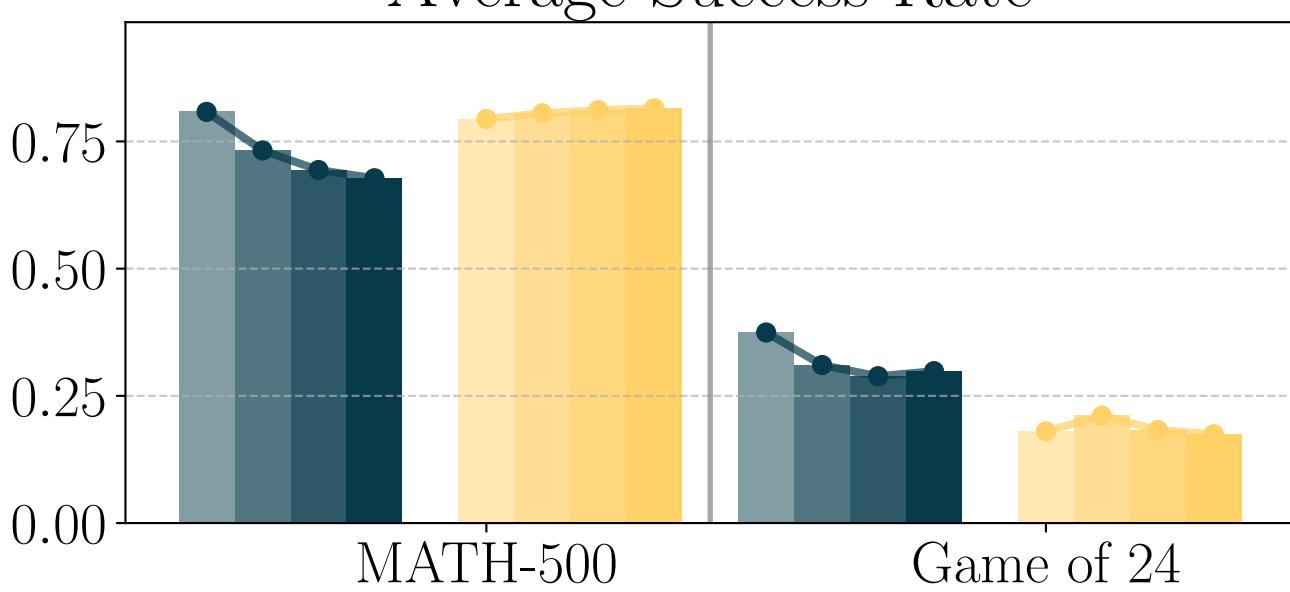
Page Solution Count



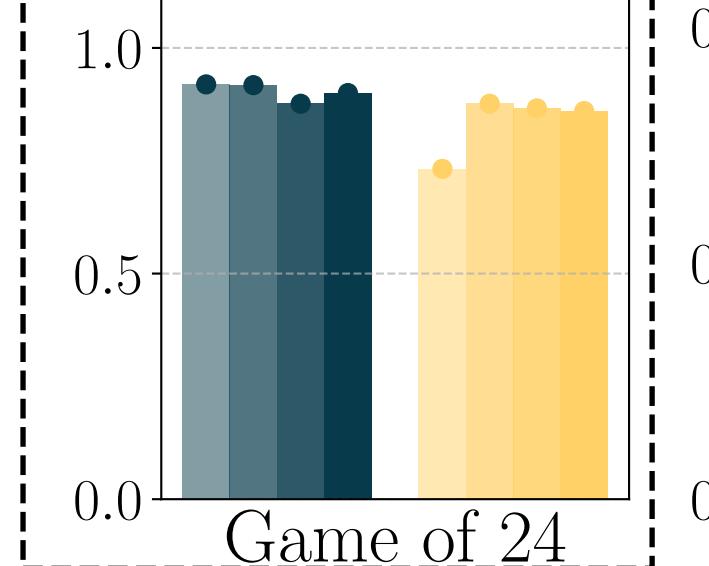
ge Jump Distance



Average Success F



age Verification Rate



The Overthinking Rate

