

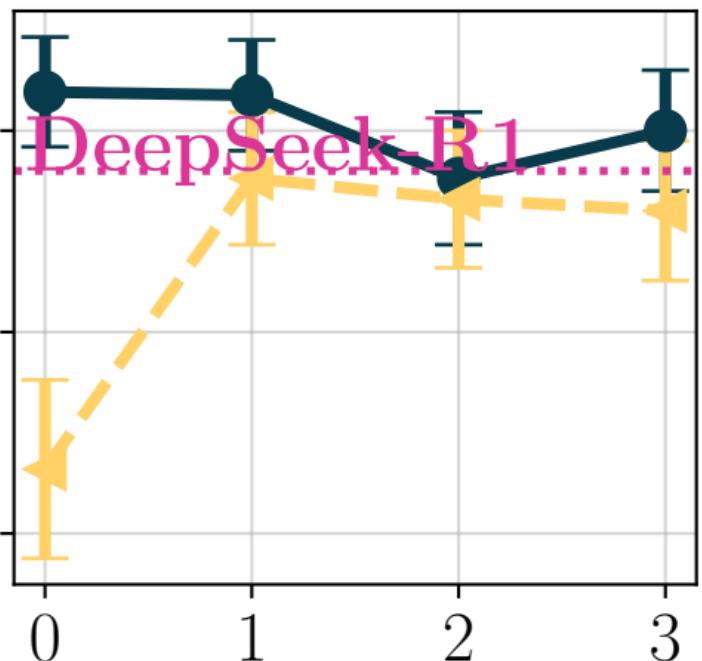


Gemini 2.0 Flash

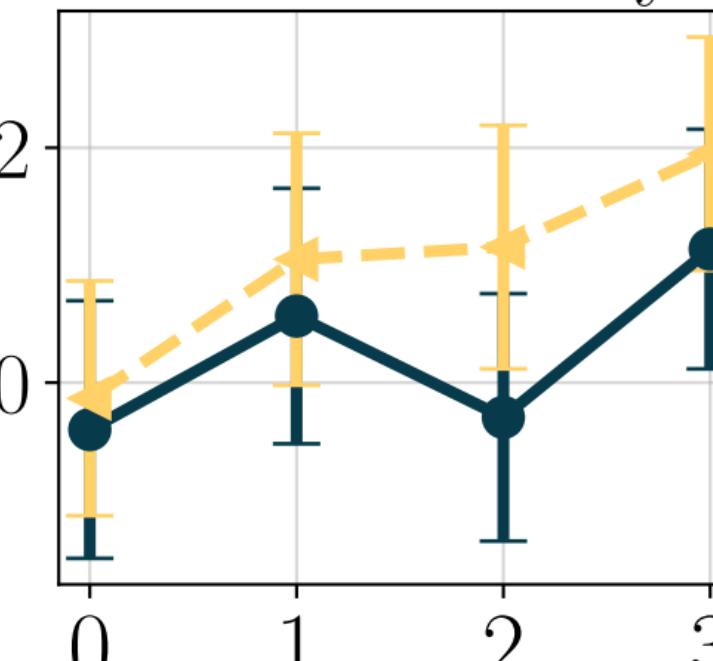


DeepSeek-V3

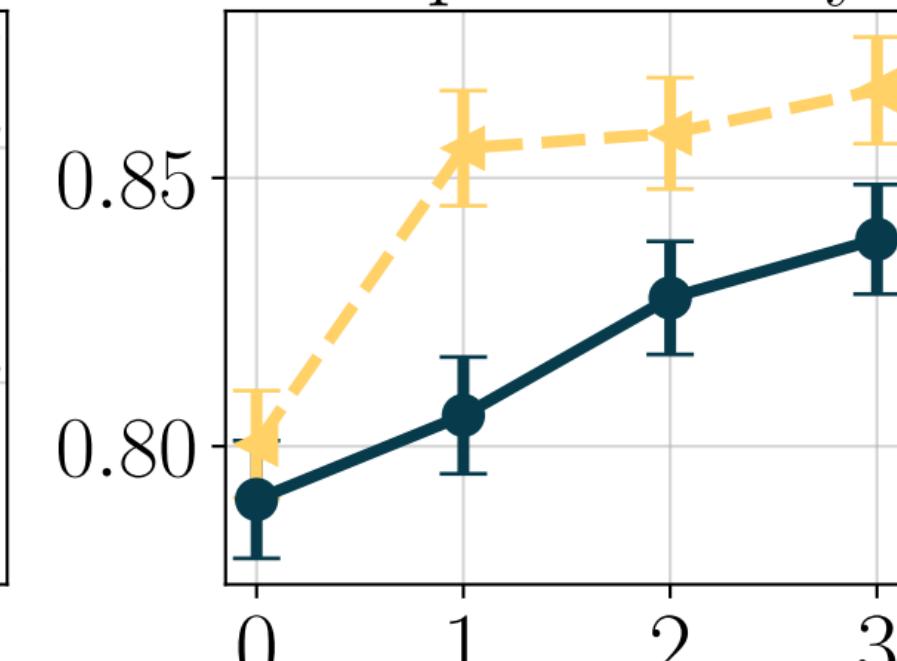
Pass@1



Tree Similarity



Jump Similarity



Number of Reasoning Examples