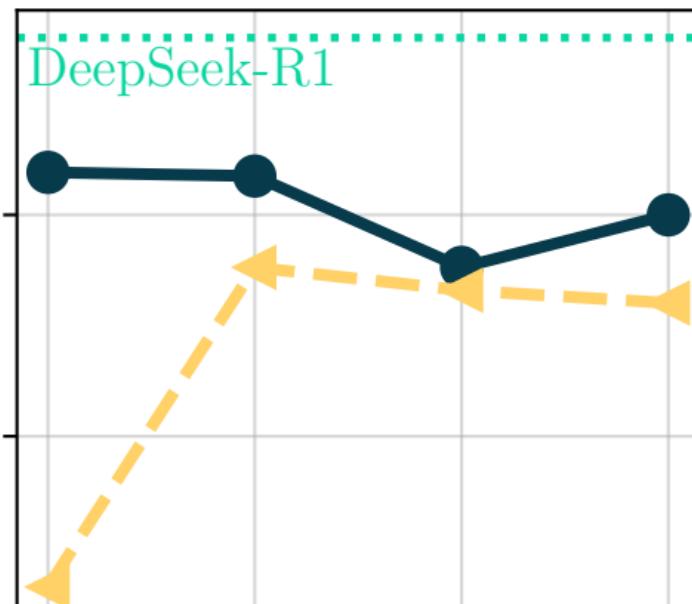
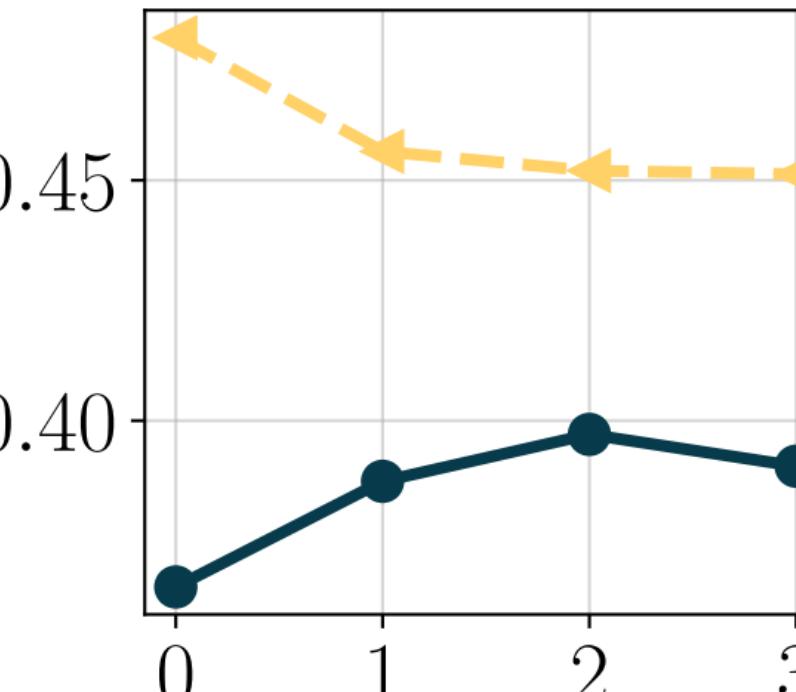


● Gemini 2.0 Flash ← DeepSeek-V3

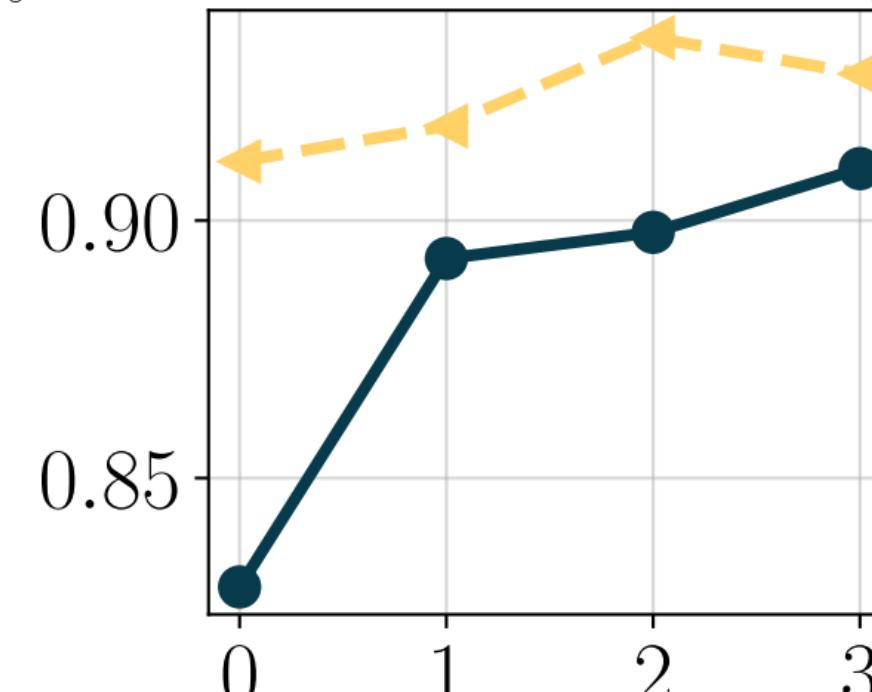
Pass@1



Plan-Structure Similarity



Action-Flow Similarity



Number of Reasoning Examples