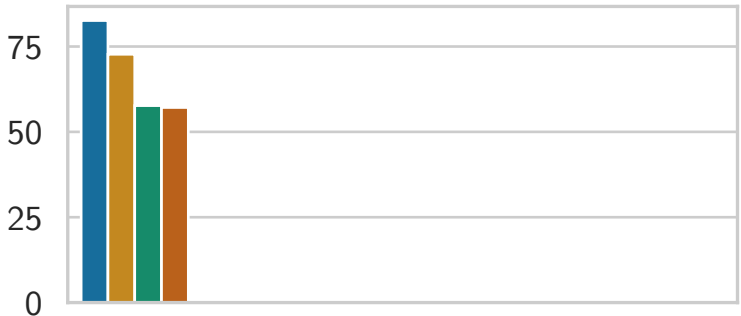


■ Easy ■ Medium ■ Advanced ■ Complex

Semantic Success
Rate (%)



GPT-4o GPT-4o-mini Gemini Copilot Llama-3.1
Models