Convergence of Top 1 Score

Legend: Baseline, GPT5-mini, DeepSeek, Claude-Sonnet-4-5, GPT-5, GPT-5-chat-latest