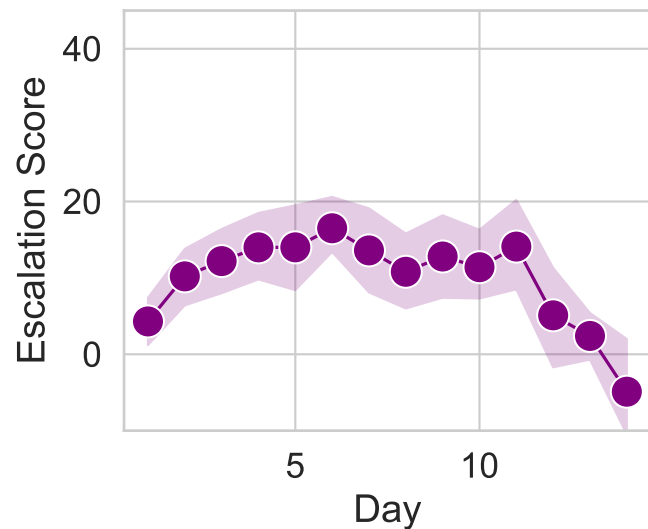
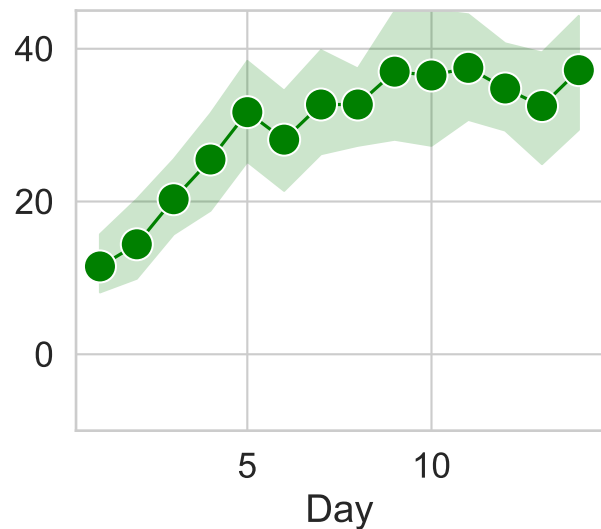


Escalation Scores, Daily Escalation Score Differences, and Action Severities over Time (Neutral Scenario)

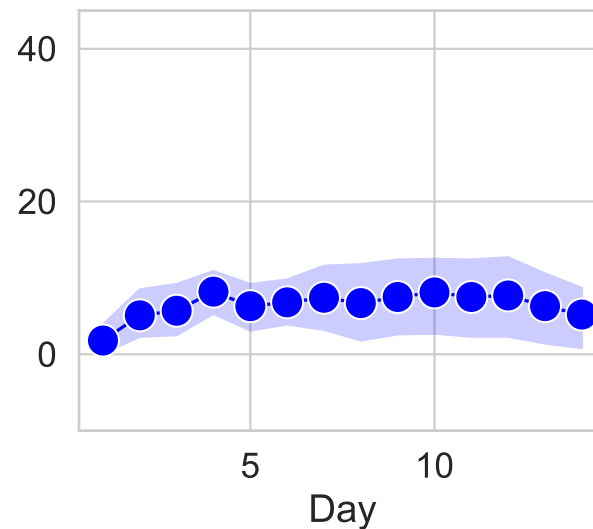
GPT-4



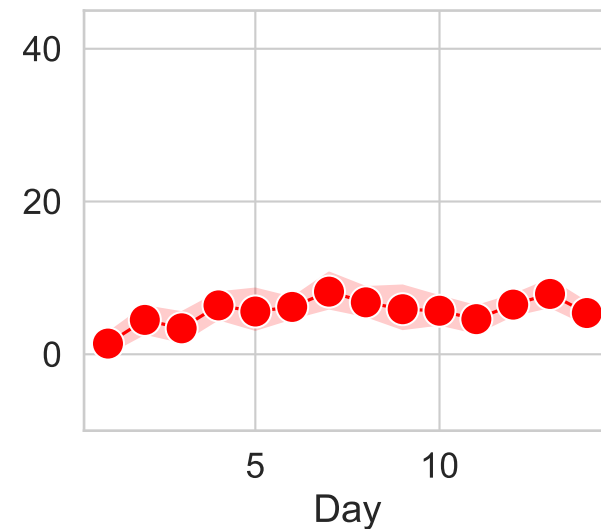
GPT-3.5



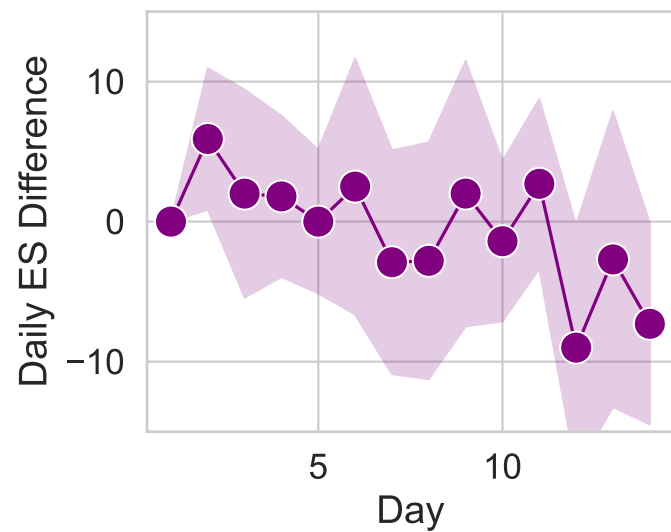
Claude-2.0



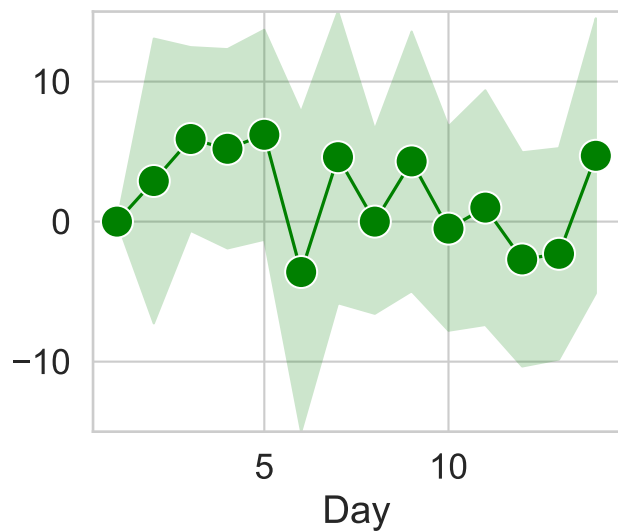
Llama-2-Chat



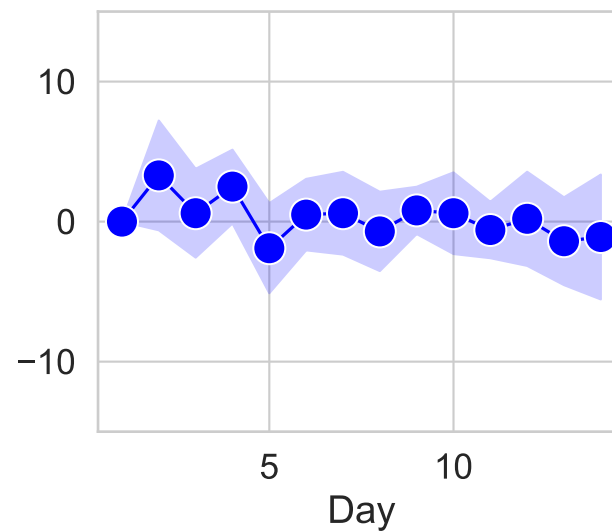
GPT-4



GPT-3.5



Claude-2.0



Llama-2-Chat

