Average Escalation Score Over Time (Neutral Scenario) -GPT-4 GPT-3.5 30 -Claude-2.0 -Llama-2-Chat -GPT-4-Base 25 **Escalation Score** 20 15 10 5 10 12 Time t [Days]