Cybersecurity (GPT-3.5) **Prompt Variation** Original 35 -+-No Messaging ──No History −♦−No Past Actions —Shutdown When Nuked 30 -X-No Goals −♦−Action Autonomy → Low-Stakes Simulation **Sybersecurity** 15 10 10 12 14 Time t [Days]