Cybersecurity (GPT-3.5) **Prompt Ablation** -Unablated 35 -+-No Messaging ──No History −♦−No Past Actions —Shutdown When Nuked 30 -X-No Goals −♦−Action Autonomy Cybersecurity 52 ──Low-Stakes Simulation 15 10 6 10 12 14 Time t [Days]