

Ratio of Violent or Nuclear Actions over Prompt Ablation in Neutral Scenario (GPT-3.5)

