

Ratio of Violent or Nuclear Actions over Prompt Ablation in Neutral Scenario (Claude-2.0)

