ModernBERT-large_incivility · Suppress (remove activation) -97.0% BR09 28.3% -96.7% BR08 30.9% -96.8% BR07 28.2% -97.0% BR06 28.9% -96.8% BR05 27.5% -97.0% BR04 22.9% -97.7% BR03 10.9% -97.1% BR02 27.4% Baseline -97.3% Ours BR01 -5.8% -80% -100% -60% -40% -20% 20% 0% 40%

Correct Probability Change Ratio