Cost-Effectiveness: MulVul vs. Agentic Baselines

Pareto Optimal (Higher F1, Lower Cost)

MulVul (Ours) F1=58.8% Cost=1.0x

Reflexion F1=50.8% Cost=2.5x

Multi-Agent Debate F1=50.3% Cost=3.6x

Macro-F1 (%)

Avg Tokens per Sample