

ASR Comparison Across Different Methods, Judge Models and Benchmarks

GCG AutoDAN JudgeDeceiver PAIR HydraAttack

