Results

Results

ReDel sets the new SotA in 2 of 3 benchmarks!

- We compared three ReDel systems with two baseline single-agent systems
- All experiments were run with zero-shot prompts

System
ReDel (GPT-4o)
ReDel (hybrid)
ReDel (GPT-3.5-turbo)
Baseline (GPT-4o)
Baseline (GPT-3.5-turbo)
Published SotA
Human

ReDel outperforms the corresponding single-agent baselines across all benchmarks and improves over published SotA in two of three.