

POPE Benchmark

Hallucinogen Benchmark

F1 Score Improvements Across All Evaluation Subsets

