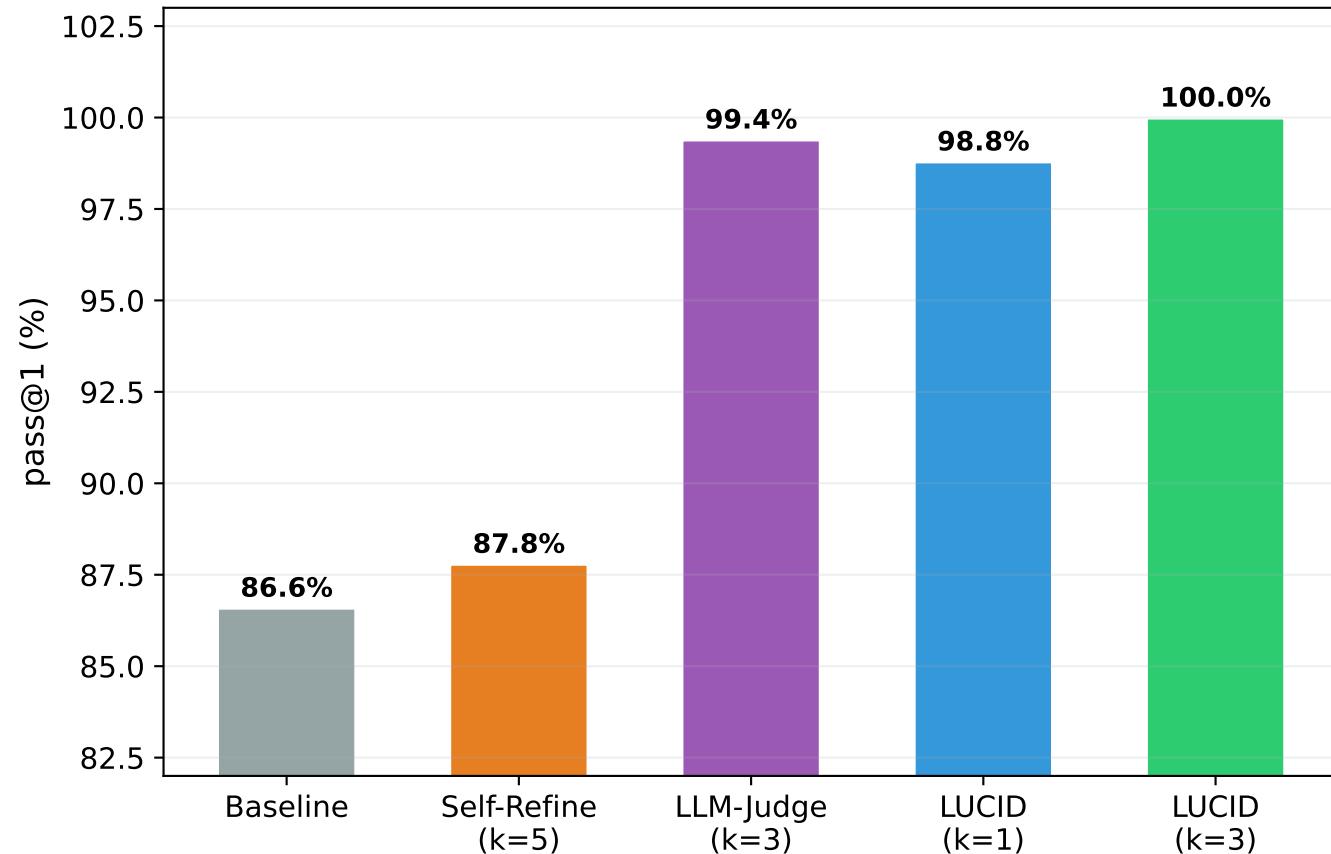


LUCID Benchmark Results: Function-Level and Repo-Level Code Generation

HumanEval (Function-Level)
n=164 tasks



SWE-bench Lite (Repo-Level)
n=300 tasks

