**SWE-bench Lite: LUCID vs Baseline
(n=300 tasks, 0 infrastructure errors)**

Resolve Rate (%)

**18.3%
(55/300)**

**25.0%
(75/300)**

+36% relative

**30.3%
(91/300)**

+65% relative

Baseline
(k=1)

LUCID
(k=1)

LUCID
(best of k=1,3)