HumanEval: Formal Verification Enables Monotonic Convergence to 100% at k=3

- pass@1 (%) vs LUCID Loop Iterations (k)
- 98.8% (162/164)
- 100.0% (164/164)
- 100.0% (164/164)
- +12.2pp
- 86.6% (142/164)
- Baseline (single pass): 86.6%
- LUCID (formal verification)