

Inference-time program synthesis and bounded execution

Phi9 · 2 October 2026 · Research demonstration

One real Codex turn generated a task-local JSON program. The expected answer was supplied. A separate Python interpreter validated and executed it.

```
[1, 2] → scale 2 → [2, 4]
→ add [1, -1] → [3, 3]
→ dot [1, 1] → scalar 6 → readout
```

15 targeted tests passed. Deterministic and CLI replay matched. Replay makes no new model request. The interpreter allows only scale, add, dot and readout, checks exact fields and finite values, and admits modeled work before operations.

What this establishes

A generated plan can run through a separate bounded execution layer and return an inspectable trace. It does not establish independent discovery, model-weight learning, unrestricted autonomy or physical meaning of the synthetic vector.

Limits

Budgets count admitted coordinate operations, logical state elements and trace records, not total process memory or provider tokens. Provider token/cost totals were not captured. Task-local configuration does not change global tool permissions.

The public source, archived program and tests are supplied for evaluation. This document is an internally checked report, not peer-reviewed research.