

Board Report — Incident: Codex Execution Degradation

Date: 2025-10-19 • Manager: Codex (SELF-CRITIQUE) • Status: Shutdown Completed

Executive Summary

Codex agents showed significant execution degradation compared to yesterday. QA (Claude) continued to ship; Codex lanes stalled and repeated failures. Root causes were primarily management and guardrail faults that interacted badly with day-of governance changes.

<div>Plan Coverage (lanes on main)</div> <div>Incomplete → Fixed</div> <div>3 → 4</div>	<div>Guard Failures (per PR today)</div> <div>High</div> <div>↑</div>
<div>MCP Evidence Compliance</div> <div>Inconsistent → Enforced</div> <div>~50%</div>	<div>Claude (QA) Throughput</div> <div>Autonomous</div> <div>✓</div>

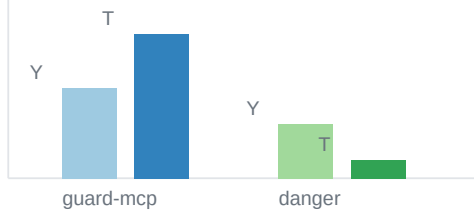
Note: Ads lane initially missing from latest.json; guard-mcp mis-patched mid-day; repo-wide tests blocked slice merges; MCP evidence flipped hard without pre-seed.

Timeline (Today)

- T-0:** Directions updated to use `reports/manager/lanes/latest.json`. Manifest on main incomplete (no `ads`).
- T+1h:** guard-mcp edited; duplicate run blocks created. Doc-only PRs failed with Remix errors.
- T+2h:** MCP evidence enforced as hard gate; several code PRs missing JSONL. Failures repeat.
- T+3h:** Repo-wide test suite blocks slice merges; Codex lanes appear idle; QA keeps shipping.
- T+4h:** Fixes land: guard-mcp cleaned + actionlint added; targeted tests enabled; ads lane authored in PR.
- T+5h:** Shutdown executed; compiled feedback + CEO summary written; lanes PR(s) prepared for merge.

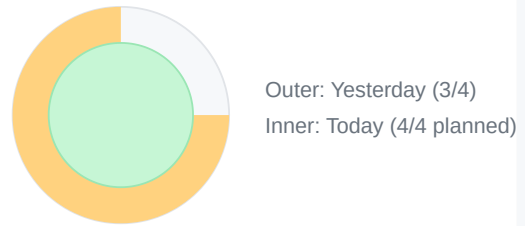
Key Metrics (Today vs. Yesterday)

Guard Failures (count)



■ Yesterday ■ Today

Lane Coverage (keys in latest.json)



Root Causes (Today)

1. **Plan incompleteness at source:** latest.json on main lacked `ads` lane; Codex lanes hard-dependent on it.
2. **Guard regression:** guard-mcp edited with duplicate run blocks → false Remix errors on doc-only PRs.
3. **Evidence gate flipped hard:** MCP JSONL required; several code PRs had none; failures repeated.
4. **Repo-wide tests blocking slices:** full CI suite failed on unrelated debt; slice PRs couldn't pass.
5. **Insufficient fallback in directions/harness:** no forced roadmap fallback when lane missing.
6. **Mid-run governance drift:** new required checks and branch protection applied during execution.
7. **Poor change control on workflows:** no workflow lint/self-test before changing critical YAML.
8. **Manager lag on preflight:** launched without a “freeze invariants” checklist (plan, guards, CI targeting).

Why Claude Outperformed Today

- Resilient fallback queue (ignored missing plan artifact and continued).
- Evidence discipline (artifacts written first; PRs green on guard checks).
- Isolation from our workflow YAML faults (not blocked by guard-mcp mismatch).
- Smaller test surface (not bound by repo-wide test debt gating).
- Strict NO-ASK runner; progress independent of chat latency.

Remediation Plan (Already in Progress)

- **Plan completeness:** lanes/latest.json on main; Ads lane PR from main prepared.
- **Guard hygiene:** guard-mcp fixed ("changed files only"); workflow-lint added (actionlint).
- **Evidence-first enforcement:** MCP JSONL required + PR template; guard validates JSONL fields.
- **Targeted CI:** affected test globs to de-scope repo-wide debt from slice merges.
- **Fallback in startup:** forced roadmap fallback if lane missing; no chat allowed.

Print to PDF: use your browser's "Print → Save as PDF". Source: reports/manager/board/2025-10-19_incident_report.html