

Degraded-Mode Playbook

Pre-defined operating procedures for when AI agent services are degraded or unavailable. Pairs with your incident post-mortem template.

TEAM	LAST TESTED	PLAYBOOK OWNER
-----	-----	-----

1. Trigger Definitions

Clear thresholds remove the "is this bad enough to act?" debate during an incident. Tune these numbers to your own baselines.

STATE	TRIGGER	WHO DECLARES	ACTION
Normal	Latency < 1.5× baseline, error rate < 2%, no provider advisory	—	Continue standard workflows
Degraded	Latency > 2× baseline for 10+ min, error rate > 5%, or provider status page showing partial outage	On-call engineer	Switch affected workflows to degraded mode; post announcement
Unavailable	Provider outage > 15 min, sustained error rate > 25%, or no successful calls in last 5 min	On-call engineer + tech lead	Activate manual mode; declare incident; notify stakeholders

2. Workflow Degraded-Mode Matrix

WORKFLOW	NORMAL (AGENT-ASSISTED)	DEGRADED (REDUCED AGENT)	MANUAL (NO AGENT)	THROUGHPUT IMPACT
Code generation	Primary model drafts, human reviews	Switch to secondary model; smaller context windows; accept shorter completions	Human writes from scratch using snippet library and internal patterns	-30% to -50%
PR / code review	Agent flags issues, human approves	Agent on priority PRs only; human handles the rest	Pair review using checklist; enforce stricter merge gates	-20% to -40%
Test writing	Agent generates scaffolds, human refines	Agent scaffolds new tests only; no bulk backfill	TDD against existing patterns; copy-adapt from sibling test files	-25% to -45%
Documentation	Agent drafts, human edits	Agent for summaries only; human drafts structure	Template-driven authoring; defer non-urgent updates	-20% to -35%
Incident triage	Agent summarises logs and suggests hypotheses	Agent for log summary only; human-led hypothesis	Runbook-driven; manual log inspection; pair on severity	-15% to -30%
Data / SQL drafting	Agent drafts queries from spec	Agent for simple SELECTs only; human writes complex joins	Query library + analyst-authored SQL	-20% to -40%
— add your own rows —				

3. Communication Templates

SLACK / TEAMS – DECLARING DEGRADED MODE

:warning: **AI services degraded** – entering **degraded mode** for [affected workflows].

Trigger: [latency / error rate / provider advisory + link]

Impact: expect [X%] slower throughput on [code gen / reviews / test writing].

What to do: follow the degraded-mode column in the playbook. Non-urgent work can continue; hold large refactors.

Owner: @[on-call] · Next update: [time]

SLACK / TEAMS – DECLARING MANUAL MODE

:rotating_light: **AI services unavailable – manual mode** now in effect.

Trigger: [provider outage / sustained errors + link]

Impact: agent-assisted workflows offline. Expect [X%] throughput reduction.

What to do: follow the manual column in the playbook. Pair up on code review. Pause any work that can safely wait.

Owner: @[on-call] · Incident channel: #inc-[id] · Next update: [time]

SLACK / TEAMS – RETURN TO NORMAL

:white_check_mark: **AI services restored** – returning to **normal mode**.

Recovery checks complete: latency, error rate, and spot-check sample all within tolerance.

Please flag any unusual agent output in the next 24h. Post-mortem to follow.

Owner: @[on-call]

4. Tabletop Exercise — The 8-Hour Outage

Goal: rehearse the degraded- and manual-mode transitions for a realistic sustained outage, without touching production. Run this once a quarter. Budget 45 minutes including debrief.

Roles

ROLE	RESPONSIBILITY
Facilitator	Reads scenario cards, keeps time, captures decisions
On-call engineer	Declares state, applies playbook, posts comms
Tech lead	Prioritises workflows, makes scope calls
Delivery lead / PM	Manages stakeholder expectations
Observers	Silent; take notes for debrief

Scenario timeline (45 minutes)

T+0 · 5 min · Brief

Facilitator sets the scene: "It's 09:30 on a release day. Primary AI provider latency has spiked to 3× baseline. The status page shows 'investigating elevated errors in the relevant region.'"

T+5 · 10 min · Card 1 — Degraded

On-call declares state, posts the degraded-mode message, identifies which workflows move. Group discusses what continues vs what pauses.

T+15 · 10 min · Card 2 — Escalation

Facilitator announces: "30 minutes in, provider confirms a regional outage. No ETA. Error rate now 80%." Group transitions to manual mode, updates comms, and the tech lead reprioritises the day.

T+25 · 10 min · Card 3 — Sustained

"Four hours in, still no recovery. A P2 bug lands in production." Group works through manual-mode incident triage without agent assistance. Who pairs with whom? What gets deferred?

T+35 · 10 min · Debrief

Structured debrief — see questions below. Capture actions with owners.

Debrief questions

- Did we declare the right state at the right time, or hesitate?
-

- Were the playbook thresholds useful, or did we argue about whether they applied?
- Which workflows were painful in manual mode? What would make them easier?
- Did our fallback tools actually work, or did we discover gaps?
- Was the comms template clear enough for people outside the room?
- What's the single change that would most improve the next run?

5. Recovery Checklist

Returning to normal mode is not just "the status page went green." Work through this before resuming full agent-assisted flows.

- Provider status page clear and stable for at least 15 minutes
- Latency and error-rate metrics back within normal thresholds
- Smoke-test the primary workflow (e.g. one code-gen request, one review) and spot-check output quality
- Compare a sample of post-recovery output against known-good baselines — watch for silent quality regressions
- Confirm no partial / truncated responses are being treated as complete
- Review any work merged during degraded or manual mode for consistency with normal standards
- Post the "return to normal" comms message
- Schedule a post-mortem within 5 working days; link this playbook from the write-up
- Capture any playbook gaps discovered during the incident and update this document

Review cadence: Review this playbook quarterly, after every real incident, and after every tabletop exercise. A playbook you haven't tested is a hypothesis, not a plan.

Degraded-Mode Playbook · Action #3 · Published on [ctco.blog](#)