

FREE PRACTITIONER PACK

Agentic AI Deployment Pack

20 checks across 5 areas. Ship the tracing with the agent — observability retrofitted after an incident is archaeology.

Pack	Agentic AI Deployment Pack
Document	Observability checklist (20 items)
Use	Production operations · launch gate
Date	2026-09

AnovaCloud — Engineering the AI-Native Enterprise. Frisco, Texas. This pack is a working instrument, not a brochure: complete it with your team, score honestly, and bring it to your discovery call. © 2026 AnovaCloud.

AnovaCloud — anovacloud.com

TRACING

☐ **Every agent run produces a full trace**

Prompt, tool calls with arguments, tool outputs, model steps, approvals — linked by a single trace ID.

☐ **Traces link to the business transaction**

Trace IDs join to the ticket, order, or workflow instance the agent was working on.

☐ **Human actions appear in the trace**

Approvals, overrides, and escalations recorded inline — the human is part of the trajectory.

☐ **Traces are queryable, not just stored**

Filter by agent, user, tool, outcome, cost — debugging without queries is guessing.

METRICS

☐ **Task success rate per task class**

The eval dimensions, measured live — success, tool correctness, groundedness proxies.

☐ **Tool error rate and retry rate**

Which tools fail, how often, and whether retries recover — per tool, per version.

☐ **Cost per task and per user**

Model + tool + infra cost attributed to tasks; budget alerts before the invoice surprises.

☐ **Latency p50/p95 per task class**

Time to first output and to completion; SLOs written for user-facing agents.

☐ **Human override and escalation rates**

How often reviewers intervene — rising overrides signal model drift or scope creep.

☐ **Approval queue health**

Queue depth, wait time, SLA breaches — the HITL tier is only as good as its queue.

ALERTING

☐ **Alerts on behavior change, not just errors**

Drift in success rate, cost spikes, novel tool arguments, unusual data access — anomaly-aware alerting.

☐ **Alert thresholds are tuned, not defaulted**

Thresholds set from baseline measurements with documented rationale; reviewed quarterly.

☐ **Alerts route to the agent's owner**

The named technical owner gets paged — not a general queue where agent alerts go to die.

RETENTION & PRIVACY

☐ **Retention policy covers traces and prompts**

How long full traces live; what is sampled vs. kept; deletion on request.

☐ **PII in traces is minimized or redacted**

Traces are a data store too — apply the same classification rules as the sources.

☐ **Sampling strategy is explicit**

What is kept at full fidelity vs. sampled, and why — decided deliberately, not by disk pressure.

DASHBOARDS & REVIEW

☐ **One dashboard per agent**

Health at a glance: success, cost, latency, overrides, incidents — the owner's morning view.

☐ **Eval-vs-production comparison**

Golden-set scores alongside live metrics; divergence investigated, not admired.

☐ **Incident traces are preserved**

Full-fidelity traces for incidents retained for postmortems and added to the eval set.

☐ **Observability itself is reviewed**

Quarterly: are we still tracing what matters? Dead metrics removed, blind spots closed.

SIGN-OFF

Technical owner: _____ Date: _____ Architect: _____ Date: _____

The observability items in this list are launch blockers — an agent you cannot see is an agent you cannot operate.