Session 5 Handout: Operating Agents in Production

FAST-13 Evaluating and Operating Agentic AI Systems · 7formations · Instructor: Wadood Chaudhary · All examples fictional

Tracing

A **trace** is one request end to end; **spans** are its steps (agent, LLM call, tool call). The OpenTelemetry GenAI semantic conventions define attributes such as gen_ai.operation.name, gen_ai.request.model, gen_ai.usage.input_tokens, and gen_ai.tool.name. In June 2026 they moved to a dedicated repository (open-telemetry/semantic-conventions-genai) and are still evolving, so check the version your platform supports.

Privacy controls (Massachusetts)

Traces can hold names with account numbers, which is personal information under G.L. c. 93H and 201 CMR 17.00. Defaults:

- Message **content capture off**; capture only the attributes you need.
- **Redact** at the collector before export.
- Set **retention** (for example 90 days), **role-based access**, encryption, and access reviews in the written information security program.
- Sandbox exception: content capture is on in class only because all data is fictional.

Offline vs. online evaluation

Use the **same graders** before and after release.

Grader	Where it runs
Golden-set regression suite	Offline: CI on every change
Cheap code checks (ID match, handoff loop)	Online: 100% of traffic
LLM judges	Online: a sample (for example 5%) plus flagged traces
Red-team suite	Offline, before release and on a schedule

Monitoring and release

- **SLOs and alerts** with thresholds (for example, ID-mismatch rate above 2% for 1 hour).
- **Release path:** CI gate → shadow → canary (for example 10%) → A/B or full rollout, with a rollback trigger and a kill switch.
- **Incident loop:** contain (roll back), review, then turn the failing traces into test cases.

Governance

Keep an eval report per release, a system card (purpose, limits, data, evaluations), and an audit trail of sign-offs. Map evaluation work to the NIST AI RMF **Measure** function. Know whether EU AI Act obligations apply to your organization.

Checklist

- ☐ Agent instrumented; agent, LLM, and tool spans visible.
- ☐ Privacy controls written: capture, redaction, retention, access.
- ☐ Online graders with rates; alerts with thresholds.
- ☐ Release gate, rollback, and named owner.
- ☐ W5 plan presented and peer reviewed.