

Who Owns AI in Production?

An Enterprise Architecture Framework for Evaluation, Governance, with Continuous Validation

Open Source Finance Forum New York

By:- Preeti Gupta

Functional Rollup to Mandated AI Roles



Traditional Application Architecture vs. AI Systems

Dimension	Traditional Application Architecture	AI Systems Architecture
Core Logic	Explicit code rules & business algorithms	Neural weights, embeddings & probabilistic models
Lifecycle Management	Continuous Integration and Continuous Deployment (CI/CD) release pipelines	Data drift monitoring, re-evaluations & fine-tuning
Quality Assurance	Unit & regression test coverage	Continuous synthetic benchmarks & real-time evaluation
Governance Perimeter	Static perimeter security & role-based access	Near real-time trace loops, guardrails & audit trails

Production Failure Modes & Risk Escalation

01. Silent Drift & Performance Decay

Model response degradation over time due to shifting input distributions, requiring automated metric thresholds.

02. Hallucination & Factual Instability

Plausible but incorrect outputs that bypass traditional data validation filters, requiring continuous semantic checks.

03. Data Leakage & Context Pollution

Unintended disclosure of sensitive business information or context contamination across multi-tenant sessions.

04. Escalation Pathway Failures

Lack of real-time circuit breakers leading to unmonitored downstream actions and compound enterprise risk.

5 Dimensions of Enterprise AI Readiness

Dim 1: Accountable Business Alignment

Defined executive ownership & Return on Investment (ROI) targets

Dim 2: Continuous Evaluation Metrics

Quantitative quality, safety & drift benchmarks

Dim 3: Architecture Integration Patterns

Standardized APIs, middleware & circuit breakers

Dim 4: Automated Risk Policy Controls

Programmatic guardrails & compliance rules

Dim 5: Operational Alerting & Remediation

Near real-time incident response & fallback actions

Continuous AI Evaluation Loop Infrastructure

PHASE 01

Pre-Production Benchmarking

1. Offline Golden Suites
2. Regression Testing
3. Pre-flight Risk Gates

Validates baseline model quality against standardized enterprise test datasets prior to release.

PHASE 02

Production Monitoring

1. Live Tracing & Logs
2. Near real-time Guardrails
3. Drift Measurement

Monitors live inference streams to detect factual errors, performance drift, and security breaches.

PHASE 03

Feedback & Remediation

1. Human Escalation
2. Dataset Enhancement
3. Model Retraining

Captures edge cases from live runtime to continuously strengthen golden benchmarks and guardrails.

Dual-Path Execution & Live Evaluation

In-Line vs. Asynchronous Evaluation Architecture

In-Line Path (Synchronous) — Applies active guardrails, token checks, and PII anonymization in real-time to prevent harm before serving user requests.

Asynchronous Evaluator Path — Runs deep LLM-as-a-judge assessments, semantic drift analysis, and compliance checks in parallel without adding user latency.

Audit Trail Logging — Persists full request-response telemetry and evaluation metrics to immutable audit logs for regulatory review.

Enterprise Architecture as Strategic Orchestrator

Core Governance Principle:

Enterprise Architecture provides the standardized frameworks and connective tissue required to operationalize governance across business, technical, and risk domains without hindering innovation.

Standardization

Establishes unified integration patterns, evaluation criteria, and toolchains across enterprise teams.

Risk Reduction

Embeds automated policy guardrails directly into runtime architectures to prevent compliance breaches.

Agility & Velocity

Enables rapid model experimentation by providing pre-certified infrastructure and clear governance gates.

AI Production Readiness Gate

Pre-Flight Release Checklist

- [] **Accountable Executive Authorization** — Formally documented sign-off confirming business case alignment, risk posture approval, and ownership designation.
- [] **Golden Benchmark Validation** — Automated verification that model accuracy, ground truth alignment, and toxicity metrics pass enterprise thresholds.
- [] **In-Line Risk Guardrails Deployed** — Inbound prompt injection defense and outbound PII/confidentiality masking active in production routing.
- [] **Near Real-Time Observability Configured** — OpenTelemetry tracing enabled with automated alerts routed to SIEM and on-call operational dashboards.
- [] **Circuit Breaker & Fallback Tested** — Validated automatic rerouting to deterministic backups or human-in-the-loop escalation upon risk breach.

Strategic Takeaways & Executive Operating Model

Key Leadership Imperative

Empowering designated AI owners requires a scalable operational framework that transitions governance from static compliance checkpoints into active, continuous system engineering.

- **Clear Ownership Boundaries:** Assign explicit operational responsibility across Accountable Executives, AI System Owners, and Governance Functions.
- **Continuous Production Evaluation:** Move beyond pre-deployment sign-offs to real-time synthetic benchmarks and automated drift detection.
- **Architectural Safeguards:** Implement in-line guardrails, OpenTelemetry tracing, and circuit breakers to enforce risk tolerance thresholds automatically.

Closing Operational Roadmap

Implementation Phasing (Days 1 – 180)

Phase 1: Immediate (Days 1–30)

Formalize AI ownership roles (AE, SO, Governance) and integrate mandatory readiness gating into release pipelines.

Phase 2: Near-Term (Days 31–90)

Establish baseline golden benchmark suites, continuous evaluation telemetry, and in-line prompt guardrails.

Phase 3: Scale (Days 91–180)

Automate closed-loop feedback, drift remediation, and policy-as-code controls across all enterprise AI deployments.

Thank you - share your feedback



Scan to rate this session