01

GOVERNANCE MOVES INTO THE PLATFORM

“Responsible AI” becomes controls.

As scrutiny rises, the winning pattern is simple: enforce governance through logging, evaluations, and policy gates in the pipeline and at runtime—so every deployment is auditable by default.

WHAT CHANGES IN 2026

Controls not docs

Logging is non-negotiable — TRACE

Centralize prompts/outputs, model versions, data lineage, and policy decisions—so incidents Are explainable, not debatable.

Evaluations become release gates — EVALS

Ship only when eval thresholds pass (hallucination, toxicity, jailbreak, privacy leakage) —and keep scores over time.

Policy enforcement moves “left” — GUARDRAILS

Access controls, allowed tools, data boundaries, and safety filters are configured once in the platform, not re-implemented per app.

Runtime monitoring is the new QA — MONITOR

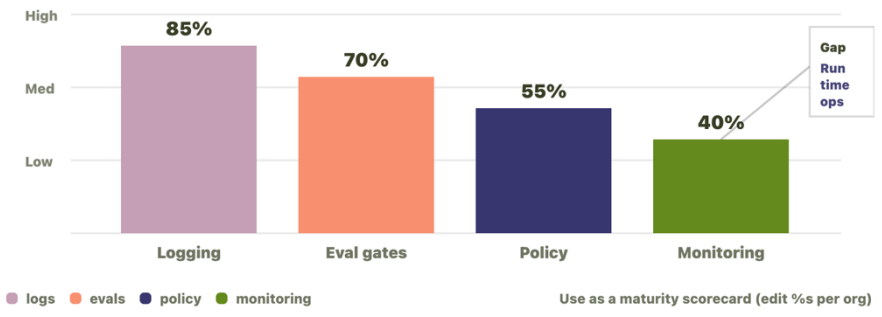
Detect drift, policy violations, and unusual tool use; trigger fallbacks, human review, or rollback when risk increases.

SIGNAL

Responsible AI = enforced controls

Exec view: the “done” definition is coverage across four controls. This turns governance into a measurable platform capability.

Target: **platform coverage** not policy docs



EXECUTIVE MOVE

Make controls a shared service

Treat responsible AI like SRE: a platform capability with measurable coverage.

Fund a **Controls Layer** with four default capabilities: **logs, eval gates, policy enforcement**, and **runtime monitoring.**

Require every AI feature to route through it— so governance scales without slowing product teams

Varun Singh
10-slide carousel Jan 2026

Sources: NIST AI RMF + Generative AI Profile