

FRAMEWORK ON ONE PAGE · VERSION 1.1

The Intent Gap

Monitoring diagnostic · May 2026

TIME TO USE 20 min review / 45 min workshop	AUDIENCE Agent owners, risk, security, and compliance teams	OUTPUT A monitored intent-gap record
--	--	---

?

For every AI system operating in production, can your organization show that actual behavior still matches the purpose, boundaries, and review conditions originally authorized? If the answer depends on assumptions, the Intent Gap is already open.

What it is

Intent Gap is the unplanned divergence between what an organization genuinely intended an AI system to do and what the system actually does in production. It is not measured at launch.

Intent Gap is not a hallucination, a generic model quality issue, or a failed launch checklist. It is not the same as Agent Sprawl or Governance Debt. Agent Sprawl is about count. Governance Debt is about accumulated missing controls.

How the gap opens

Operational authority exceeds intent	The system is authorized for a narrow purpose, but its permissions, tool access, or execution path allow actions beyond that purpose.
Use changes after deployment	The system starts in one workflow, then becomes part of a higher-stakes decision path.
Monitoring measures activity, not intent	Logs can show what happened. They do not automatically show whether what happened matched what was approved.

A worked example

Situation	A generic enterprise notices that a support agent has begun including recommendations outside its approved summarization purpose.
Where it lands	Intent gap requires review
Evidence	Observed outputs include action recommendations that were not included in the approved intent record.
Gap found	The trigger threshold and disposition owner are not attached to the monitoring record.
Next action	Open an intent-gap record, review the evidence, and decide whether to constrain, suspend, or re-authorize the agent.
Output	A documented mismatch with a named review path.

The one action it produces

A monitored intent-gap record

What good looks like

Every production AI system has an authorization record that states purpose, authorized actions, explicit prohibitions, expected outputs, data access, accountable owner, review cadence, and escalation path.

What it cannot establish

This object can identify a documented mismatch between purpose and observed behavior. It cannot explain every causal factor or prove that monitoring captured all relevant behavior.

CITE AS