# The AI Plumber

## A Governance-First Framework for Regulated Agentic AI

Koen Van Lysebetten · March 2026

## 1 Executive Summary

AI agents are being deployed faster than the governance infrastructure around them can evolve. This creates a compliance debt that high-risk, regulated environments — banking, healthcare, public sector — cannot afford. The AI Plumber framework closes that gap by making governance the first architectural layer, not an afterthought. Every agent action is attributable and logged, every policy envelope is defined before deployment, and every kill switch is tested before it is needed. In a world of the EU AI Act and sectoral regulators, governance-first is not a constraint on velocity; it is the only architecture that survives a regulator, a board, and a production incident simultaneously.

**RULE: Use agentic AI only when you can log, attribute, and reverse every contextual judgment in an audit-ready format.**

## 2 The Problem: Infrastructure Before Intelligence

Most AI deployments start with model selection and end with a governance retrofit. In regulated environments, this sequence is unworkable. Three failure modes dominate:

| | | |
|---|---|---|
| No audit trail | AI model cites stale third-party data | Regulatory liability (GDPR Art.9) & reputational damage |
| No rollback | Schema injection corrupts live CMS | Production incident & manual reconciliation |
| No kill switch | Agent continues publishing after threshold breach | Compliance violation & platform ban |

*Table 1: Common failure modes in under-governed AI deployments*

These are not edge cases. They are systemic risks that become production incidents in the absence of a governance-first architecture, mapping directly onto logging, human oversight, and risk management obligations now imposed on high-risk AI systems under the EU AI Act.

# 3 The AI Plumber Framework

Four foundational patterns operationalize governance as infrastructure:

## Pattern 1: Constrained Agent Identities

Each agent operates under a narrowly scoped service account with explicit resource and action boundaries. No agent inherits human user privileges. This reduces blast radius and directly supports data protection mandates in sectoral frameworks (PDPL/SAMA, GDPR). Cryptographic verification at every service boundary.

## Pattern 2: Attributable Actions

Every agent decision is logged with full input context, reasoning trace, and output action. This creates a forensically auditable and reversible decision trail, satisfying record-keeping requirements for high-risk AI systems. 100% reversible decision trails — no black-box AI.

## Pattern 3: Human-in-the-Loop Gates

High-stakes actions (financial commitments, legal publishing, policy changes) require explicit human approval before execution. Human oversight is architecturally enforced — the workflow mechanically pauses and awaits a human authorization token — not merely mentioned in a policy PDF.

## Pattern 4: Kill Threshold Monitoring

Continuous telemetry tracks agent behavior against predefined safety thresholds: velocity spikes, cost overruns, error rate breaches. Threshold violations trigger automatic suspension and human escalation. This operationalizes ex-post monitoring mandated by the EU AI Act for high-risk deployers.

# 4 Agentic AI vs. Traditional Automation

When to use agentic AI is as important as how. The decision matrix below prevents misapplication of an architecture that requires fundamentally different governance:

| | | |
|---|---|---|
| State space | Finite, enumerable | Unbounded, contextual |
| Failure modes | Fully specified | Emergent |
| Governance model | Change management | Live policy envelope |
| Audit requirement | IT change log | Decision + reasoning trace |
| Regulatory fit | Product safety / IT change management | EU AI Act, sectoral guidelines (SAMA, RIZIV) |

*Table 2: Decision matrix — when to use agentic AI*

## 5 Three-Phase Deployment Model

Governance gates scale with automation scope. Each phase unlocks the next only when the prior governance layer is operational and audited:

| | | |
|---|---|---|
| Phase 1 | €50K | Risk register · EU AI Act high-risk classification · GDPR Art.9 data classification map · Read-only scope |
| Phase 2 | €500K | Policy envelope · Kill thresholds · Human gates for all write actions · Rollback capability |
| Phase 3 | €5M+ | Multi-client policy layer · Agent confidence network · Full orchestration scope |

*Table 3: Governance scales with automation scope*

## 6 Production Proof Points

The framework is not theoretical. The following cases represent governance-first AI deployed under some of the world's strictest regulators:

| | | | |
|---|---|---|---|
| Najm Insurance (Saudi Arabia) | Insurance claims | SAMA compliance · PDPL data protection · 40 cities | 6,000+ daily cases · zero-tolerance misclassification · hybrid cloud + edge |
| De Lijn (Belgium) | Public transport AI roadmap | EU AI Act · C-suite governance · 5,000+ FTE impact | 129% projected ROI · 3-year roadmap · board approved |
| US Restaurant Intelligence | Operational intelligence | Cost efficiency · real-time pipeline · audit observability | 200-person workflow → 3 agents · 1 month → 10 min · ~90% cost reduction |

| | | | |
|---|---|---|---|
| NAMA Museum (India) | Cultural heritage | Sovereign data residency · archival integrity · public accountability | €10M+ program · 180+ projectors · 99.9% SLA · read-only + audit trail |

*Table 4: Governance-first AI in production*

# 7 Conclusion: Plumbing Is the Moat

The technology itself is rarely the risk. The system around it is. Manufacturing dependency, regulatory naivety, no audit trail — these are the failure modes that kill enterprise AI. The AI Plumber framework ensures that every agent action is attributable, every policy envelope is defined before deployment, and every kill switch is tested before it is needed.

**Governance-first is not a constraint on velocity.**
**It is the only architecture that survives a regulator, a board, and a production incident simultaneously.**