

A GOVERNANCE-FIRST APPROACH TO REGULATED AGENTIC AI

The Why Behind *the Tech*

Governance. Observability. No magic, just infrastructure.

"Every architectural choice deserves a paper trail. Not to cover your back — to set your team free."

KOEN VAN LYSEBETTEN · MARCH 2026 · AIPLUMBER.DEV

EU AI ACT

SAMA

GDPR ART. 9

GOVERNANCE-FIRST

AGENTIC AI

REGULATED ENTERPRISE

Koen Van Lysebetten

AI Architect & Governance Advisor · March 2026

INSURANCE

PUBLIC SECTOR

BANKING

HEALTHCARE

Governance as the **First Layer**

Most AI deployments start with model selection and end with a governance retrofit. In regulated environments — banking, healthcare, insurance, public sector — this sequence creates compliance debt that no organization can afford under the EU AI Act, SAMA, GDPR Article 9, or sectoral frameworks.

The AI Plumber framework reverses this: governance becomes 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's needed.

CORE THESIS

In regulated AI, the moat is not the model — it's the **governance layer.**

4

FOUNDATIONAL
PATTERNS

3

DEPLOYMENT
PHASES

6

GOVERNANCE
PIPES

90d

TO GOVERNED
PRODUCTION

Infrastructure Before **Intelligence**

Three failure modes appear repeatedly across enterprise AI deployments in regulated industries. They are not edge cases — they are systemic risks that become production incidents without governance-first architecture. Each maps directly to logging, human oversight, and risk management obligations now imposed on high-risk AI systems under the EU AI Act.

TABLE 1 — COMMON FAILURE MODES IN UNDER-GOVERNED AI DEPLOYMENTS

FAILURE MODE	REAL-WORLD SIGNAL	CONSEQUENCE
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

× TRADITIONAL SEQUENCE (BROKEN)

- 01 Select foundation model
- 02 Build application layer
- 03 Run pilot with limited scope
- 04 Scale to production
- 05 Retrofit governance when regulator asks

Problem: By step 5, you have compliance debt, no audit trail, and a system that cannot prove its decisions in a regulatory review.

✓ AI PLUMBER SEQUENCE (FIXED)

- 01 Define governance requirements and risk classification
- 02 Build control plane: logging, attribution, rollback, kill switches
- 03 Implement constrained agent identities
- 04 Deploy agents within policy envelope
- 05 Scale with continuous telemetry and human gates

Result: Every agent action is auditable, reversible, and attributable from day one. Governance infrastructure scales with automation scope.

Four Foundational Patterns

The AI Plumber framework is built on four patterns that, taken together, create a governance layer capable of satisfying the most demanding regulatory environments — from the EU AI Act to SAMA, GDPR, and beyond.

Pattern 01

CONSTRAINED AGENT IDENTITIES

Problem: Agents that inherit human privileges create unlimited blast radius and regulatory liability.

Solution: Each agent operates under a narrowly scoped service account with explicit resource and action boundaries.

Pattern 02

ATTRIBUTABLE ACTIONS

Problem: AI decisions without reasoning trails are black boxes that fail audit requirements.

Solution: Every agent decision is logged with full input context, reasoning trace, and output action.

Pattern 03

HUMAN-IN-THE-LOOP GATES

Problem: Fully autonomous agents in high-stakes scenarios create unacceptable regulatory and operational risk.

Solution: High-stakes actions require explicit human approval before execution — architecturally enforced.

Pattern 04

KILL THRESHOLD MONITORING

Problem: Agents without real-time safety monitoring can spiral into costly or dangerous behavior before humans notice.

Solution: Continuous telemetry tracks agent behavior against predefined safety thresholds with automated suspension.



PATTERN 01

CONSTRAINED AGENT IDENTITIES

No agent inherits human user privileges. Service accounts are scoped to the minimum required permissions, with cryptographic verification at every service boundary. Read-only access is the default; write access requires explicit justification and is logged at the policy level.



PATTERN 02

ATTRIBUTABLE ACTIONS

Every agent decision is logged with: timestamp and agent ID; input context (sanitized for PII); reasoning trace or model output; action taken; confidence score; and decision rationale. This creates a 100% reversible decision trail — if an agent publishes incorrect content, you can trace back to the exact input that triggered the error, review the reasoning, and reverse the action.

✓ REGULATORY ALIGNMENT: EU AI ACT ARTICLE 12 · GDPR ARTICLE 22 · EX-POST MONITORING



PATTERN 03

HUMAN-IN-THE-LOOP GATES

The workflow mechanically pauses and awaits a human authorization token before executing any high-stakes action. The agent generates a proposed action, a notification is sent to a human approver, the system waits for an approval token, the action executes only after explicit authorization, and the full approval chain is logged.

TABLE 2 — HIGH-STAKES ACTIONS REQUIRING HUMAN APPROVAL GATES

CATEGORY	EXAMPLES
Financial commitments	Transactions over defined threshold
Legal document publishing	Contracts, compliance filings
Policy changes	Actions affecting user data
Schema modifications	Production database changes
Customer-facing communications	Regulated industry outbound messaging



PATTERN 04

KILL THRESHOLD MONITORING

Monitored thresholds include: velocity (actions per minute/hour exceeding baseline); cost (API spend above budget ceiling); error rate (failed actions or rejected outputs above tolerance); confidence decay (model confidence scores trending below acceptable range); and policy violations (attempts to access restricted resources).

The automated response cascade is: threshold breach detected → agent automatically suspended → human escalation notification sent → incident log created with full context → system awaits manual review and restart authorization.

Decision Matrix

TABLE 3 – WHEN TO USE AGENTIC AI

DIMENSION	AGENTIC AI	TRADITIONAL AUTOMATION
State space	Unbounded, contextual	Finite, enumerable
Failure modes	Emergent	Fully specified
Governance model	Live policy envelope	IT change management
Audit requirement	Decision + reasoning trace	IT change log
Regulatory fit	EU AI ACT · SAMA · RIZIV	Product safety / IT change guidelines



CRITICAL RULE

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

DEPLOYMENT MODEL

Three-Phase Deployment

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

TABLE 4 – GOVERNANCE SCALES WITH AUTOMATION SCOPE

PHASE	ARR BAND	GOVERNANCE GATE
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

PHASE 1

€50K

ARR

READ-ONLY INTELLIGENCE

Objective: Prove value with zero operational risk. Agents operate with read-only access to data sources, analysis and reporting only — no write actions, no external API calls.

→ Validated use case, initial audit trail, no production risk.

PHASE 2

CONTROLLED AUTONOMY

€500K

ARR

Objective: Enable agent write actions with full human oversight and rollback. Write actions to internal systems, API integrations, content generation and publishing — all within constrained agent identities.

→ Production-grade automation with regulatory-ready governance layer.

PHASE 3

€5M+

ARR

ORCHESTRATED INTELLIGENCE

Objective: Multi-agent orchestration with enterprise-scale governance. Multi-agent workflows with handoffs, cross-client policy layer, agent confidence network, full orchestration scope across systems.

→ Enterprise AI factory with governance infrastructure that scales.

PRODUCTION PROOF

Governance-First in Production

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

TABLE 5 – GOVERNANCE-FIRST AI IN PRODUCTION

CLIENT	DOMAIN	CONSTRAINT	RESULT
Najm Insurance (Saudi Arabia)	Insurance	SAMA compliance · PDPL claims data protection · hybrid cloud + edge	6,000+ daily cases · zero-tolerance misclassification · 40 cities
De Lijn (Belgium)	Public transport	EU AI Act · C-suite governance · AI roadmap board approved	129% projected ROI · 3-year roadmap · 5,000+ FTE impact

CLIENT	DOMAIN	CONSTRAINT	RESULT
US Restaurant Intelligence	Operational	Cost efficiency · real-time audit · observability pipeline	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 · audit trail



INSURANCE · SAUDI ARABIA

NAJM INSURANCE

6,000+ daily insurance claims across 40 cities with zero-tolerance policy on misclassification in a hybrid cloud + edge environment. Constrained agent identities for each claims processor, full audit trail for every classification decision, human approval gates for claims above SAR threshold.

SAMA COMPLIANT

PDPL CERTIFIED

40 CITIES

6,000+ DAILY CASES



PUBLIC TRANSPORT · BELGIUM

DE LIJN

AI transformation roadmap for Belgium's largest public transport operator (5,000+ FTE). Complete EU AI Act risk assessment, board-approved 3-year AI roadmap with governance gates, human oversight architecture for customer-facing AI.

129% ROI

BOARD APPROVED

EU AI ACT

5,000+ FTE



OPERATIONAL INTELLIGENCE · USA

US RESTAURANT INTELLIGENCE

Three specialized agents (order fulfillment, invoice processing, compliance check) replaced a 200-person manual workflow. Real-time telemetry pipeline with kill thresholds, continuous audit trail for all agent actions.

~90% COST REDUCTION

3 AGENTS

10 MIN PROCESSING

900K+ LOCATIONS



CULTURAL HERITAGE · INDIA

NAMA MUSEUM

€10M+ program with 180+ projectors and 99.9% SLA. Sovereign data residency compliance, archival integrity, public accountability with read-only access and full audit trail.

€10M+ PROGRAM

99.9% SLA

180+ PROJECTORS

SOVEREIGN DATA



FROM THE FIELD

"We replaced 200 operators with 3 AI agents — the first decision was not the model, it was what happens at 3am when it's wrong."

PRODUCTION CLIENT · US RESTAURANT INTELLIGENCE

IMPLEMENTATION

12-Week Roadmap

The following 12-week roadmap provides a structured path from initial risk assessment to full production deployment with regulatory-ready governance infrastructure.

WEEKS 1-2

RISK ASSESSMENT & CLASSIFICATION

- Conduct EU AI Act risk classification
- Map GDPR Article 9 data exposure
- Document sectoral compliance requirements
- Create initial risk register
- Define high-stakes actions requiring human approval

Deliverable: Risk assessment document and compliance requirements matrix

WEEKS 3-4

CONTROL PLANE ARCHITECTURE

- Design logging and attribution infrastructure
- Implement constrained agent identity system
- Build human approval workflow
- Deploy telemetry and threshold monitoring
- Create rollback mechanism

Deliverable: Governance infrastructure operational in staging environment

WEEKS 5-6

POLICY ENVELOPE DEFINITION

- Define allowed actions per agent identity
- Set kill thresholds (velocity, cost, error rate, confidence)
- Map human approval gates to high-stakes actions
- Document policy envelope in audit-ready format
- Test kill switches and rollback procedures

Deliverable: Policy envelope documented and tested

WEEKS 7-8

PHASE 1 DEPLOYMENT (READ-ONLY)

- Deploy agents with read-only access
- Monitor telemetry and logging
- Validate attribution and audit trail
- Review governance gates with compliance team
- Measure baseline performance

Deliverable: Production deployment with zero operational risk, validated audit trail

WEEKS 9-12

PHASE 2 DEPLOYMENT (CONTROLLED AUTONOMY)

- Enable write actions with human gates
- Deploy kill threshold monitoring
- Test rollback procedures in production
- Conduct internal governance audit
- Scale to additional use cases

Deliverable: Full automation with regulatory-ready governance layer

MONTH 4+

PHASE 3 SCALING (ORCHESTRATED INTELLIGENCE)

- Multi-agent orchestration with handoffs
- Cross-client policy isolation (if applicable)
- Agent confidence network deployment
- Board-level governance reporting
- Continuous compliance monitoring

Deliverable: Enterprise AI factory with governance infrastructure that scales

CONCLUSION

Plumbing is the Moat

Governance-First is the **Only** **Architecture** **That Survives.**

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's 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.

NEXT STEPS

→ Download the complete AI Plumber whitepaper at aiplumber.dev

→ Schedule a governance readiness assessment for your organization

→ Join the AI Plumber newsletter for implementation case studies and framework updates

→ Book a workshop: 2-day AI governance sprint for enterprise teams

ABOUT THE AUTHOR



Koen Van Lysebetten

AI ARCHITECT & GOVERNANCE ADVISOR

AI Architect and Governance Advisor specializing in regulated agentic AI for banking, healthcare, insurance, and public sector organizations. He has led AI transformations at De Lijn (Belgium public transport), Najm Insurance (Saudi Arabia), and multiple enterprise clients across Europe and the Middle East.

He writes about AI governance, growth architecture, and enterprise AI implementation at koenvanlysebetten.substack.com and advises organizations through DevGap and Digital Dali Labs.

KOEN@AIPLUMBER.DEV AIPLUMBER.DEV [LINKEDIN](#) [SUBSTACK](#)

REFERENCES

- [1] European Parliament. (2024). *Regulation (EU) 2024/1689 on Artificial Intelligence (AI Act)*. Official Journal of the European Union.
- [2] Saudi Central Bank (SAMA). (2023). *Regulatory Framework for Artificial Intelligence in Financial Services*. www.sama.gov.sa
- [3] European Commission. (2023). *Guidelines on High-Risk AI Systems under the AI Act*. digital-strategy.ec.europa.eu
- [5] National Institute for Health and Disability Insurance (RIZIV). (2024). *AI Guidelines for Healthcare Providers*. Belgium.
- [4] Van Lysebetten, K. (2025). *The Vertical Collapse: Why AI Agents Are Betting on Vertical*. Digital Dali Growth Mentor. koenvanlysebetten.substack.com
- [5] National Institute for Health and Disability Insurance (RIZIV). (2024). *AI Guidelines for*