The AI Plumber Framework

A Governance-First Approach to Regulated Agentic AI

Koen Van Lysebetten • March 2026

Executive Summary

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.

This whitepaper presents the four foundational patterns, the three-phase deployment model, production proof points from regulated industries, and a practical implementation roadmap.

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

The Problem: Infrastructure Before Intelligence

Three Critical Failure Modes

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

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

These aren't edge cases—they're 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.

The Traditional Deployment Sequence (Broken)

1. Select foundation model
2. Build application layer
3. Run pilot with limited scope
4. Scale to production
5. 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.

The AI Plumber Sequence (Fixed)

1. Define governance requirements and risk classification
2. Build control plane: logging, attribution, rollback, kill switches
3. Implement constrained agent identities
4. Deploy agents within policy envelope
5. 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.

The Framework: Four Foundational Patterns

Pattern 1: 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.

Figure 1: Every architectural choice deserves a paper trail—not to cover your back, but to set your team free

Implementation:

- No agent inherits human user privileges
- Service accounts scoped to minimum required permissions
- Cryptographic verification at every service boundary
- Read-only access as default; write access requires explicit justification

Regulatory alignment: Directly supports data protection mandates in sectoral frameworks (PDPL/SAMA, GDPR), reduces exposure under EU AI Act Article 9 (risk management systems).

Pattern 2: 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.

What gets logged:

- Timestamp and agent ID
- Input context (sanitized for PII)
- Reasoning trace or model output
- Action taken
- Confidence score
- Decision rationale

Forensic auditability: Create 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: Satisfies record-keeping requirements for high-risk AI systems (EU AI Act Article 12), supports GDPR Article 22 (automated decision-making), enables ex-post monitoring.

Pattern 3: 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. Human oversight is architecturally enforced—the workflow mechanically pauses and awaits a human authorization token.

Figure 2: The AI Plumber: Plumbing as the moat in regulated AI

High-stakes actions that require human gates:

- Financial commitments over defined threshold
- Legal document publishing
- Policy changes affecting user data
- Schema modifications in production databases
- Customer-facing communications (regulated industries)

Implementation: Agent generates proposed action → Notification sent to human approver → System waits for approval token → Action executes only after explicit authorization → Full approval chain logged.

Regulatory alignment: Operationalizes EU AI Act Article 14 (human oversight), supports financial sector requirements (SAMA, MiFID II), enables accountability under GDPR Article 5.

Pattern 4: Kill Threshold Monitoring

Problem: Agents that operate 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. Threshold violations trigger automatic suspension and human escalation.

Monitored thresholds:

- **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
- **Policy violations:** Attempts to access restricted resources

Automated response cascade:

1. Threshold breach detected → Agent automatically suspended
2. Human escalation notification sent
3. Incident log created with full context
4. System awaits manual review and restart authorization

Regulatory alignment: Operationalizes ex-post monitoring mandated by the EU AI Act for high-risk deployers (Article 61), supports continuous oversight requirements in healthcare (RIZIV) and finance (SAMA).

When to Use Agentic AI vs. Traditional Automation

Not every problem needs agentic AI. The wrong architectural choice creates unnecessary governance overhead or fragile automation.

Dimension	Traditional Automation	Agentic AI
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

Use traditional automation when:

- Decision tree is fully mappable
- All edge cases can be anticipated
- Workflow is deterministic
- Standard IT governance suffices

Use agentic AI when:

- Context is unbounded and evolving
- Human-like judgment is required
- Failure modes are emergent
- You can build the governance layer first

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

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	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

Table 3: Governance scales with automation scope

Phase 1: Read-Only Intelligence (€50K ARR)

Objective: Prove value with zero operational risk.

Agent scope:

- Read-only access to data sources
- Analysis and reporting only
- No write actions, no external API calls

Governance requirements:

- Complete risk register
- EU AI Act high-risk classification assessment
- GDPR Article 9 data classification map
- Basic logging and attribution

Outcome: 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.

Agent scope:

- Write actions to internal systems
- API integrations with external services
- Content generation and publishing
- Constrained agent identities deployed

Governance requirements:

- Policy envelope defining allowed actions
- Kill thresholds with automated suspension
- Human approval gates for high-stakes actions
- Full rollback capability tested in staging
- Continuous telemetry dashboard

Outcome: Production-grade automation with regulatory-ready governance layer.

Phase 3: Orchestrated Intelligence (€5M+ ARR)

Objective: Multi-agent orchestration with enterprise-scale governance.

Agent scope:

- Multi-agent workflows with handoffs
- Cross-client policy layer (for agencies/platforms)
- Agent confidence network (agents evaluate each other)
- Full orchestration scope across systems

Governance requirements:

- Multi-client policy isolation
- Agent-to-agent attribution chains
- Distributed kill switch coordination
- Real-time compliance monitoring

- Board-level governance reporting

Outcome: Enterprise AI factory with governance infrastructure that scales with automation scope.

Figure 3: Building AI Factories: The five-layer cake that powers your intelligent future

---

Production Proof Points

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

Client	Domain	Constraint	Result
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

Case Deep Dive: Najm Insurance (Saudi Arabia)

Context: Insurance claims processing under SAMA (Saudi Central Bank) compliance and PDPL (Personal Data Protection Law).

Challenge: 6,000+ daily insurance claims across 40 cities with zero-tolerance policy on misclassification in a hybrid cloud + edge environment.

AI Plumber implementation:

- Constrained agent identities for each claims processor
- Full audit trail for every classification decision
- Human approval gates for claims above SAR threshold
- Kill switches for error rate anomalies
- Sovereign data residency compliance

Outcome: Production deployment meeting SAMA compliance requirements, PDPL data protection standards, and operational SLAs across distributed infrastructure.

Case Deep Dive: De Lijn (Belgium Public Transport)

Context: AI transformation roadmap for Belgium's largest public transport operator (5,000+ FTE).

Challenge: EU AI Act compliance for high-risk AI systems, C-suite governance approval, public sector accountability.

AI Plumber implementation:

- Complete EU AI Act risk assessment and classification
- Board-approved 3-year AI roadmap with governance gates
- Human oversight architecture for customer-facing AI
- Continuous monitoring and ex-post review framework

Outcome: 129% projected ROI with full regulatory compliance, board approval, and clear governance accountability structure.

Case Deep Dive: US Restaurant Intelligence

Context: Operational intelligence for multi-location restaurant chain.

Challenge: Replace 200-person manual workflow while maintaining audit observability and cost efficiency.

AI Plumber implementation:

- Three specialized agents (order fulfillment, invoice processing, compliance check)
- Real-time telemetry pipeline with kill thresholds
- Continuous audit trail for all agent actions
- Cost monitoring with automatic suspension triggers

Outcome: 200-person workflow reduced to 3 agents, processing time from 1 month to 10 minutes, approximately 90% cost reduction with full auditability.

Figure 4: 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

---

Implementation Roadmap

Week 1-2: Risk Assessment and Classification

- Conduct EU AI Act risk classification
- Map GDPR Article 9 data exposure
- Document sectoral compliance requirements (SAMA, RIZIV, MiFID II, etc.)
- Create initial risk register
- Define high-stakes actions requiring human approval

Deliverable: Risk assessment document and compliance requirements matrix.

Week 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.

Week 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.

Week 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.

Week 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.

Wireframe: Governance Dashboard

Real-Time Monitoring View

- **Agent Status Panel:** Active agents, suspended agents, agents awaiting approval
- **Threshold Monitoring:** Velocity, cost, error rate, confidence score (real-time graphs)
- **Kill Switch Status:** Active thresholds, recent violations, suspension history
- **Human Approval Queue:** Pending approvals, average approval time, approval rate
- **Audit Trail Access:** Search by agent ID, timestamp, action type, decision rationale

---

Policy Envelope Configuration

- **Agent Identity Management:** Service accounts, permission scopes, resource boundaries
- **Action Allowlist:** Permitted actions per agent, required approval gates
- **Threshold Configuration:** Set velocity limits, cost ceilings, error tolerances
- **Rollback Procedures:** Automated rollback triggers, manual rollback interface

- **Compliance Reporting:** EU AI Act checklist, sectoral requirements, audit readiness score

---

Incident Response Interface

- **Active Incidents:** Agent suspensions, threshold breaches, policy violations
- **Incident Timeline:** Full event log with reasoning trace
- **Manual Override:** Emergency stop, manual restart, policy exception
- **Root Cause Analysis:** Input context, reasoning trace, action taken, failure point
- **Remediation Actions:** Rollback execution, notification history, escalation path

---

Visual Taglines for Each Framework Component

For "The Why Behind the Tech" (Image 1)

1. Every architectural choice deserves a paper trail—not to cover your back, to set your team free.
2. Build the pipes before the magic.
3. Governance isn't a constraint. It's the moat.
4. The first question isn't "Can AI do this?" It's "What happens when it's wrong at 3am?"
5. In regulated AI, documentation is infrastructure.

For "The AI Plumber: Plumbing as the Moat" (Image 2)

1. Plumbing is the moat in regulated AI.
2. The technology is not the risk. The system around it is.
3. Before autonomy comes attribution, rollback, and control.
4. Don't start with the agent. Start with the failure path.
5. In regulated environments, the control plane is the product.

For "Building AI Factories" (Image 3)

1. Layer 5 collapses without Layer 1. Build from the bedrock up.
2. Smart rewards require successful, informed AI built on stable foundations.
3. The factory is only as intelligent as its governance infrastructure.
4. Wobbly foundations cause total collapse—build the datacenter bedrock first.
5. Enterprise AI is not a demo. It's a five-layer stack with governance pipes at every level.

For "We Replaced 200 Operators with 3 AI Agents" (Image 4)

1. 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.
2. Fragile AI chaos is what you get without governance pipes, policy layers, and shutdown paths.
3. Controlled compliant production is not optional in regulated environments—it's the only architecture that survives a board, a regulator, and a production incident simultaneously.
4. Visualization is first-class infrastructure. You can't govern what you can't see.
5. The continuous audit trail is not a nice-to-have. It's the legal requirement.

Distribution Strategy: Where to Ship

Option 1: aiplumber.ai (Recommended)

Why this works:

- Positions AI Plumber as a standalone methodology/framework (not just Koen's consulting)
- Opens path to productization (governance platform, audit dashboard, certification program)
- Creates category-level IP that can be licensed or partnered

- Easier to pitch to enterprises ("We use the AI Plumber framework" vs. "We use Koen's approach")
- Separates thought leadership from personal brand

Site structure:

- **Landing page:** Hero visual, core thesis, whitepaper download
- **/framework:** Four foundational patterns with interactive wireframes
- **/case-studies:** Najm, De Lijn, US Restaurant, NAMA deep dives
- **/roadmap:** 12-week implementation guide
- **/resources:** Blog, newsletter signup, webinar registration
- **/contact:** Enterprise inquiry form, workshop scheduling

Lead capture: Email gate on whitepaper download, newsletter for framework updates, webinar series on each pattern.

Long-term play: AI Plumber becomes a recognized methodology (like Jobs-to-be-Done, Design Sprint, or Service Design Thinking), you become the authority, enterprises pay for certification/training/implementation.

Option 2: koenvanlysebetten.com (Personal brand play)

Why this works:

- Consolidates all thought leadership under your name
- Reinforces you as the creator and authority
- Easier to cross-promote with Substack, LinkedIn, speaking gigs
- Lower friction to ship (one domain to maintain)

Site structure:

- **/work:** BrandMind, DevGap, consulting clients
- **/frameworks:** AI Plumber, Growth Architecture, LLMO
- **/writing:** Substack integration, long-form essays
- **/speaking:** Speaking topics, past events, booking form
- **/contact:** Advisory inquiries, workshop scheduling

Lead capture: Unified email list for all frameworks, single CRM flow.

Long-term play: Koen Van Lysebetten becomes the recognized expert, AI Plumber is one of several frameworks under your name, easier to pivot or expand.

Option 3: Hybrid Approach (Best of both)

- **aiplumber.ai:** Dedicated site for the framework (whitepaper, case studies, roadmap, certification)
- **koenvanlysebetten.com:** Personal hub linking to AI Plumber and other ventures (BrandMind, DevGap, Substack)
- Cross-link strategically: "Created by Koen Van Lysebetten" on aiplumber.ai, "Explore AI Plumber" on koenvanlysebetten.com

Why this works: You build IP that can live beyond personal consulting while maintaining personal brand equity. AI Plumber can become a licensed framework, training program, or software product without cannibalizing your personal consulting brand.

Recommended Launch Sequence

Week 1: Domain and Site Setup

- Register aiplumber.ai (recommended) or decide on koenvanlysebetten.com/ai-plumber
- Set up landing page with hero visual, core thesis, email capture form
- Create whitepaper PDF (this document with professional design treatment)
- Integrate email capture (ConvertKit, Mailchimp, or HubSpot)

Week 2: Content and SEO

- Publish complete framework documentation on site
- Create case study pages for Najm, De Lijn, US Restaurant, NAMA
- Optimize for search: "governance-first AI," "EU AI Act compliance framework," "regulated agentic AI"
- Set up blog with first 3 posts (one per week)

Week 3: Distribution Campaign

- LinkedIn post series: One post per framework pattern (4 posts total)
- Substack feature: "Introducing AI Plumber" with link to full whitepaper
- Email blast to existing network: Personal intro + whitepaper link
- Submit to Product Hunt, Hacker News, AI-focused communities

---

Week 4: Speaking and Partnerships

- Pitch webinar to enterprise AI communities (AWS, GCP, Databricks user groups)
- Submit talk proposal to AI/governance conferences (ReWork AI Summit, AI World Forum)
- Reach out to complementary vendors (Dust, Claude, governance platforms) for co-marketing
- Schedule intro calls with existing clients for case study interviews and testimonials

---

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'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.ai
- 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 is an 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.

Contact: koen@aiplumber.ai | LinkedIn: koenvanlysebetten

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*. https://www.sama.gov.sa

[3] European Commission. (2023). *Guidelines on High-Risk AI Systems under the AI Act*. https://digital-strategy.ec.europa.eu

[4] Van Lysebetten, K. (2025). The Vertical Collapse: Why AI Agents Are Betting on Vertical. *Digital Dali Growth Mentor*. https://koenvanlysebetten.substack.com

[5] National Institute for Health and Disability Insurance (RIZIV). (2024). *AI Guidelines for Healthcare Providers*. Belgium.