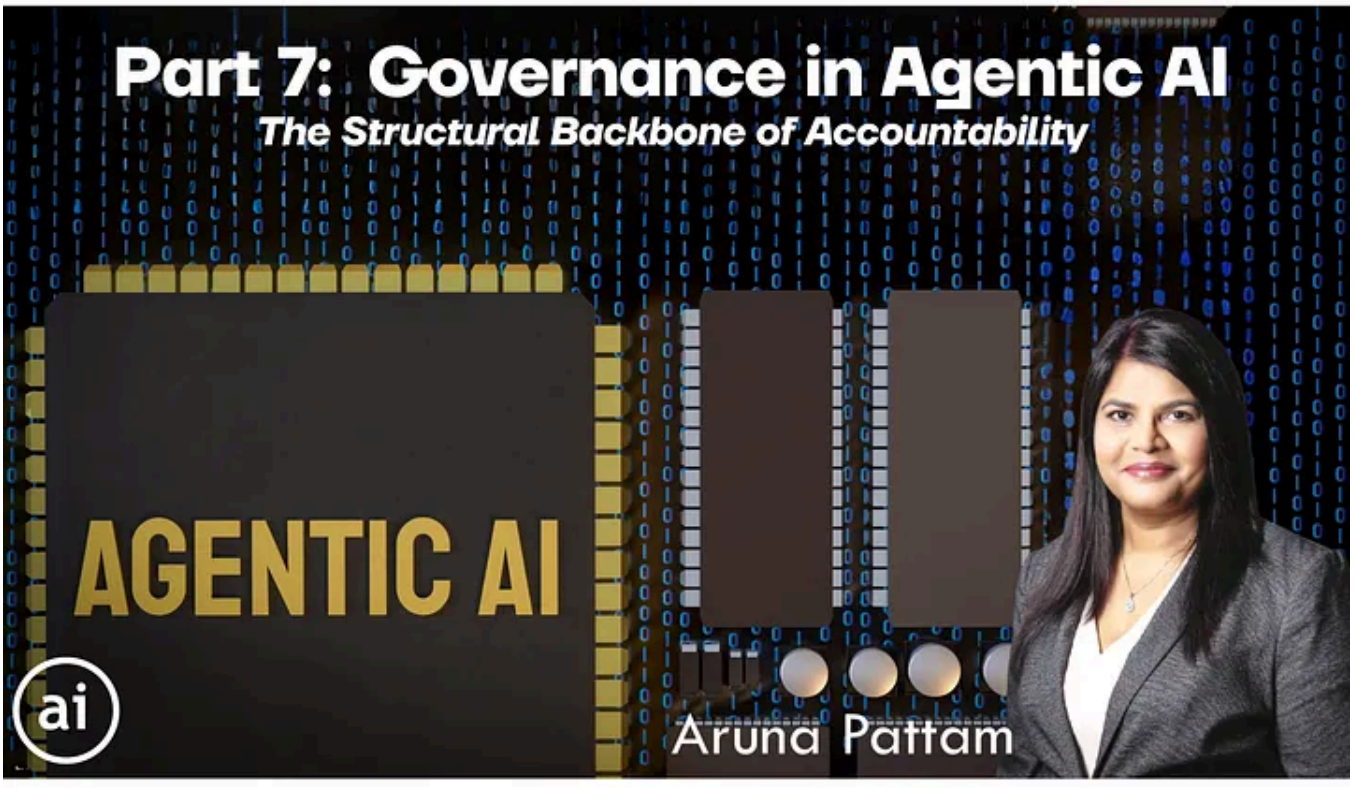
Agentic AI: Part 7— Governance in Agentic AI

8 min read · 6 days ago

Aruna Pattam Follow



In *Part 6* of this series, we explored how **observability** transforms trust from an abstract belief into measurable assurance allowing organisations to see *why* their autonomous agents act the way they do.

But visibility alone isn't enough. Knowing *what's happening* doesn't ensure that what's happening is *right*.

As Agentic AI systems grow in autonomy reasoning, delegating, and making business-critical decisions the question becomes sharper:

If *trust* is the foundation and *observability* is the scaffolding, then **governance** is the structure that keeps everything upright. It ensures that agents operate safely, ethically, and within regulatory boundaries, even as they act at machine speed.

Without governance, Agentic AI becomes a collection of intelligent silos powerful, fast, but unaccountable.

With governance, it becomes an **ecosystem of controlled autonomy**, where every action is authenticated, traceable, and aligned with enterprise intent.

In insurance, an industry built on accountability governance is not optional; it's existential.

Why Governance Matters in the Age of Autonomy

Traditional software systems follow human instructions and their behaviour is predictable. But agentic systems are different. They decide, act, and even adapt based on context.

This shift changes the very nature of accountability.

When dozens of AI agents interact across underwriting, pricing, claims, and compliance, we need new rules of engagement:

- Who authorised each decision?
- What boundaries constrain each agent?
- How do we prevent a chain reaction when one agent fails?
- Can regulators audit every step of the AI-driven decision?

Without governance, autonomy quickly becomes anarchy.

Governance in Agentic AI gives two key kinds of assurance:

1. **Structural Control**: making sure every agent has clear roles, permissions, and accountability.

2. **Behavioural Alignment**: ensuring agents make decisions that stay true to human values, ethics, and regulations.

Zero-Trust Principles for Agent Governance

The first step toward effective governance is adopting a **Zero-Trust mindset**.

In traditional IT, Zero-Trust means “never trust, always verify.” The same principle now applies to autonomous agents.

No agent, no matter how reliable, should be trusted implicitly.

Every interaction must be verified. Every access must be authorised. Every action must leave an immutable footprint.

1. Decentralised Identities (DIDs)

Every agent must have its own verifiable identity, cryptographically secured and registered within an organisational ledger.

In insurance, this means an underwriting agent, pricing agent, and compliance agent each carry unique digital identities. If a dispute arises such as an incorrect quote or an unfair decision, investigators can trace exactly *which agent* acted, *when*, and *under whose authority*.

DIDs create **digital accountability** at scale, the backbone of trust in distributed ecosystems.

2. Verifiable Credentials (VCs)

Just as humans carry job titles and access badges, agents carry verifiable credentials that define their authority.

A pricing agent's credential might allow it to compute premiums but not approve final quotes. A compliance agent's credential might allow it to enforce rules but not modify them.

Credentials ensure **role clarity**. No agent acts beyond its scope, even if it tries.

3. Fine-Grained Access Control

Governance thrives on precision. Instead of broad, role-based access, Agentic AI requires **attribute-based controls**, permissions that adapt dynamically to context.

For example:

· A triage agent can view customer submission data but not financials.

· A pricing agent can access actuarial tables but not personally identifiable information (PII).

This principle ensures that autonomy doesn't come at the cost of security.

4. Context-Aware Policies

Static governance policies fail in dynamic systems.
Context-aware policies allow agents to adapt based on situational parameters such as transaction value, risk level, or regulatory jurisdiction.

If a commercial property policy exceeds a certain threshold, the system automatically triggers human review. If an AI model drifts beyond acceptable fairness variance, workflows pause until compliance approves.

Context transforms governance from rigid rule-following to **intelligent oversight**.

5. Immutable Audit Trails

Accountability ends where transparency stops.

Every agent decision, credential exchange, and policy check must generate immutable logs permanent, time-stamped, and tamper-proof.

For insurers, this means a regulator can reconstruct a decision chain months later understanding every step without ambiguity.

Audit trails don't just protect against non-compliance; they **build confidence** among stakeholders, regulators, and customers alike.

Governance Models for Multi-Agent Systems

Governance is not one-size-fits-all.

The right model depends on business size, regulatory complexity, and risk appetite.

Let's examine the four primary models and why most insurers will gravitate toward a **hybrid approach**.

1. Centralised Governance

All decisions and policies flow through a single authority typically a compliance or AI ethics board.

This model provides strong control and clear accountability, but limits agility.

Best suited for: highly regulated insurers and early-stage Agentic AI deployments.

2. Federated Governance

Each business unit (e.g., life, property, commercial lines) governs its own agents under enterprise-wide standards.

It balances oversight with autonomy but risks uneven policy enforcement.

Best suited for: large insurers operating across multiple geographies.

3. Decentralised Governance

Agents operate autonomously within distributed environments, following predefined rules and self-enforcing compliance.

This model boosts scalability and adaptability but can make unified oversight harder.

Best suited for: innovation ecosystems requiring high flexibility and independent agent coordination.

4. Hybrid Governance (Dominant Model for Insurance)

Combines centralised oversight (for compliance and ethics) with distributed autonomy (for operational agents).

Functional agents across underwriting, pricing, and claims operate autonomously, while a central governance layer continuously audits outputs for compliance and ethics.

Best suited for: enterprise-grade insurance ecosystems where efficiency must coexist with regulatory confidence.

Insurance Example: Underwriting Governance in Action

Imagine a commercial underwriting ecosystem built on governed autonomy:

- Each AI agent has a **DID** (identity).
- The **Pricing Agent** is credentialed only to calculate premiums — not to approve them.
- The **Compliance Agent** is credentialed only to verify rules and fairness thresholds.
- **High-value risks** automatically escalate to human review.
- Every decision, correction, and escalation is immutably logged.

If a regulator later questions why a quote was declined or a rate was adjusted, the insurer can replay the full chain of reasoning, every data source, every rule check, every human intervention.

This is governance in action: speed without recklessness, autonomy with accountability.

The Measurement Gap in Agentic AI

Even with strong governance, most organisations struggle with measurement.

They evaluate agents on accuracy or efficiency, but not on human trust, fairness, or long-term stability.

This imbalance creates what I call **The Measurement Gap**, a disconnect between technical success and real-world adoption.

A truly effective governance framework for Agentic AI requires evaluation that goes far beyond technical metrics. Accuracy and efficiency matter, but they are not enough to guarantee trust, adoption, or long-term reliability.

A balanced approach measures performance across four key dimensions: **Technical, Human-Centered, Temporal, and Contextual** to ensure AI systems remain accountable and valuable in real-world environments.

1. Technical Axis

This dimension focuses on traditional performance indicators such as accuracy, latency, and throughput.

High technical scores mean an AI system operates efficiently, producing quick and precise results.

For instance, an underwriting agent may calculate premiums in seconds with minimal error.

However, even perfect technical precision cannot ensure adoption if the system fails to meet human or ethical expectations.

2. Human-Centered Axis

The human element determines whether people actually trust and use the system. This axis measures interpretability, usability, and alignment with real-world workflows.

A technically flawless pricing model that brokers find confusing or opaque will be rejected in practice.

The key question is: can users understand and rely on the agent's reasoning?

For example, providing clear explanations for premium adjustments increases trust and collaboration.

3. Temporal Axis

AI performance must endure over time. This axis evaluates stability, resistance to drift, and adaptability to new data or regulations.

A claims assessment agent that performs well at launch but degrades after months of evolving customer patterns represents poor temporal governance.

Regular retraining, auditing, and policy updates ensure sustained reliability.

4. Contextual Axis

Finally, governance must account for broader business, ethical, and regulatory contexts.

In insurance, this means ensuring fairness across portfolios, compliance with evolving laws, and measurable ROI. A contextually aware AI doesn't just perform tasks, it operates responsibly within the organisation's values and market conditions.

Together, these four axes create a governance framework that is **balanced, dynamic, and human-aligned**, a living system built for both performance and accountability.

Governance in Practice: Beyond Policy to Precision

Governance in Agentic AI isn't about paperwork or bureaucracy, it's about precision by design. It must be embedded into every layer of the AI lifecycle to ensure that autonomy remains accountable and aligned with human intent.

Before deployment, governance establishes the foundation by defining clear rules, ethical boundaries, and escalation paths. Each agent is assigned roles and permissions, ensuring it knows *what* it can do, *when*, and *under whose authority*.

During execution, governance takes an active role through continuous, real-time monitoring. It verifies that agents adhere to compliance policies, data access rules, and fairness thresholds while maintaining operational efficiency. Anomalies or deviations trigger alerts, allowing immediate corrective action before risks escalate.

After decisions are made, governance provides reflection through audits, post-deployment reviews, and iterative improvement. Logs, outcomes, and user feedback are analysed to strengthen both technical and ethical alignment.

In this way, governance becomes more than a safety net, it becomes the *heartbeat of accountability*. It links technical architecture with ethical architecture, ensuring that every autonomous action is explainable, compliant, and in service of the organisation's values.

Conclusion

Agentic AI challenges us to redefine responsibility.

When an autonomous agent makes a decision that impacts a customer's life or livelihood, accountability must still trace back to a human name.

The question for every enterprise is not just "Can our agents act autonomously?" but:

"Can we prove, at any moment, that their autonomy remains aligned with our intent?"

That's the essence of governance, not restricting intelligence, but directing it.

Connecting the Pillars

As we conclude the exploration of *Trust, Observability, and Governance*, the architecture of responsible autonomy becomes clear:

For insurers, these are not academic ideals, they are the foundation of the next-generation enterprise: one that is **trustworthy, transparent, and governable by design.**

Agentic AI doesn't just require intelligent systems; it requires *intelligent stewardship*.

Follow the links below for previous part:

Agentic AI: Part 6 — End-to-End Observability in Agentic AI

The Operational Scaffolding of Trust

arunapattam.medium.com

Agentic AI: Part 5— Trust in Agentic AI

The Cornerstone of Adoption

arunapattam.medium.com

Agentic AI: Part 4 — Agentic AI Frameworks

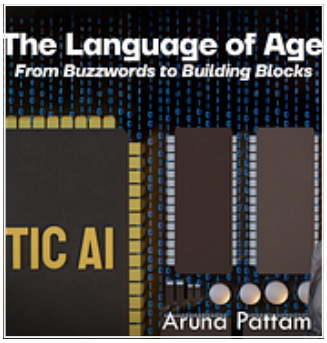
Introduction to Agentic AI Frameworks

arunapattam.medium.com

Agentic AI: Part 3- Key Terms You Should Know

In Part 2, we explored how AI agents have progressed from simple chatbots to intelligent systems that can plan, act...

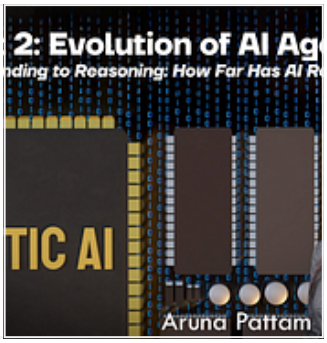
arunapattam.medium.com



Agentic AI: Part 2–Evolution of AI Agents

In Part 1, we explored what Agentic AI is, how it differs from traditional AI, and why it matters more than ever. If...

arunapattam.medium.com



Agentic AI: Part 1-Introduction

Imagine having an AI that doesn't just wait for your instructions but actually understands your goals, makes plans, and...

arunapattam.medium.com

AI | Agentic Ai | Ai Governance | Innovation

Follow

Written by Aruna Pattam

709 followers · 33 following

I head AI Platforms at Zurich, driving GenAI & Agentic AI adoption, building scalable frameworks, and championing ethical, diverse AI.

Responses (1)

Bgerby

What are your thoughts?

Jeff Urbanczyk
6 days ago

Thanks Aruna for another great write up. Can you share some more practical examples (including architectures) of how best to implement governance that remains of high quality over time? Thanks.

Reply

More from Aruna Pattam

Aruna Pattam

Agents in Generative AI: A Comprehensive Overview

The world of artificial intelligence (AI) is advancing at breakneck speed, and one of the most groundbreaking developments is the emergence...

Jul 31, 2024 121 4

Aruna Pattam

Agentic AI: Part 6—End-to-End Observability in Agentic AI

The Operational Scaffolding of Trust

Oct 6 63

Aruna Pattam

Agentic AI: Part 1-Introduction

Imagine having an AI that doesn't just wait for your instructions but actually understands your goals, makes plans, and gets things done...

Jun 14 60 3

Aruna Pattam

Agentic AI: Part 2–Evolution of AI Agents

In Part 1, we explored what Agentic AI is, how it differs from traditional AI, and why it matters more than ever. If you haven't read it...

Jun 29 16

See all from Aruna Pattam

Recommended from Medium

Aruna Pattam

Agentic AI: Part 1-Introduction

Imagine having an AI that doesn't just wait for your instructions but actually understands your goals, makes plans, and gets things done...

Jun 14 60 3

Aman Raghuvanshi

Agentic AI #5—AI Workflows vs AI Agents: What's the Real Difference?

Understand the differences between AI workflows and AI agents. Learn when to use each, with real-world examples, code, and a hybrid system.

Jul 1 118 1

Aakash Gupta

AI Just Made Product Discovery 10x Faster (Here's the Exact Process)

Product discovery used to take weeks.

Oct 16 59 9

In Agentic AI & GenAI Revolution by Yi Zhou

The New AI Mandate: Why Every CIO's 2026 Strategy Must Include Agentic Engineering

Discover why 2026 marks the shift from GenAI to Agentic AI—and how CIOs can drive ROI through the new discipline of Agentic Engineering.

6d ago 5 2

Dr Nicolas Figay

Knowledge Graphs and Ontologies: Beyond the Dictionary Fallacy

Most knowledge graph practitioners treat ontologies as sophisticated dictionaries—structured vocabularies and entity hierarchies...

Oct 18 104 2

Steve Jones

AI Orchestration—slap a sticker on the workflow

“Orchestration” isn’t new, and there is a reason it isn’t “the” answer

Oct 17 111 2

See more recommendations