Introduction

Problems with Current Governance Models

Currently, many organizations function without a formal AI governance structure or strategic plan, creating potential compliance and operational risks. Those that have implemented AI policies and procedures are currently not designed to oversee the rapidly evolving and highly interactive systems advancing into autonomous technologies. As a result, current governance models operate as fragmented layers, each with its own structural weaknesses. These weaknesses compound across the ecosystem, creating inconsistent oversight, limited traceability, and significant compliance risks, particularly for small and mid-sized enterprises (SMEs) that lack the resources to employ dedicated AI governance and ethical compliance teams.

The shortcomings fall into three major domains:

- Collective (Global) Governance
- Relational (Human-AI Interaction) Governance
- Internal (Technical/Model) Governance

Each domain underperforms in different ways, but together they form a pattern of misalignment and operational fragility that impedes safe and auditable AI deployment and accelerated use for the betterment of the individual, an organization, and society at large.

Collective (Global Governance)

Collective Governance refers to regulators, standards bodies, and cross-organizational frameworks responsible for defining risk classifications, safety controls, and compliance expectations. In practice, these mechanisms often fail to operate at the speed or granularity required for modern AI systems moving toward agentic autonomy.

The current state of Collective Governance is characterized by:

- Slow Response Cycles: Regulatory updates, standards revisions, and compliance guidance often lag technological capability by months or years. This delay leaves organizations without clear direction during critical adoption phases.
- Siloed Frameworks: National, sectoral, and organizational frameworks frequently operate independently. EU AI Act classifications, NIST AI RMF guidance, ISO/IEC 42001 controls, state-level regulations (e.g. TRAIGA), OECD AI principles rarely align in structure, content, and intent. Organizations must navigate overlapping and contradictory requirements in a continuously transforming sector.
- Reactive Postures: Polices are often shaped reactively after a failure or noncompliance forced review of AI deployment practices. Regulators and

companies act when incidents have already occurred, creating cyclical waves of compliance changes rather than creating stable, proactive governance structures.

The result is that collective governance provides high-level constraints but not the operational infrastructure required for effective continuous oversight. Organizations are left to interpret and implement requirements with limited guidance, creating a wide variation in execution and capability.

Relational (Human-AI Interaction) Governance

Relational Governance covers human-AI supervision, decision review, auditing processes, and oversight of interactive or agentic systems. This layer is the most operationally fragile in SMEs and the least standardized across industries.

Significant risk factors include:

- Inconsistent Human Oversight: Human evaluators vary widely in expertise, risk awareness, and evaluation criteria. The result is inconsistent decision quality and unpredictable reinforcement signals to AI systems.
- Lack of Standardization: Currently, there are no shared protocols for the following which results in incompatible or incomplete oversight of AI applications
    - How to define organizational structure and individual accountability for AI use
    - How to audit human-AI interactions
    - How to label model behaviors
    - How to evaluate agent actions
    - How to document human-AI decision paths
- Lack of Transparency: Most relational oversight produces no verifiable audit trail, making it impossible to reconstruct past model behavior or demonstrate compliance. If they exist, many relational frameworks rely on
    - Ad hoc judgement calls to audit
    - Ephemeral interactions
    - Both human and AI decisions undocumented in real time
- Human-Dependency and Risk of Burden: As with all auditing frameworks, risk increases when evaluators are undertrained, teams are understaffed, use of the system surges, and complexity of the process accelerates.

The result is that the relational governance layer becomes the weakest layer of modern AI oversight due to the increase in direct human-AI interactions directly in opposition with lack of auditable structures and consistent individual training. It lacks repeatability, scalability, and evidentiary tracing, creating the biggest gap in regulatory compliance.

Internal (Technical/Model) Governance

Internal governance covers the embedded safety mechanisms, technical controls, model guardrails, and integral auditability structures within AI systems themselves.

Internal governance models consistently underperform because they are:

- Optional: Many organizations treat internal safety, logging, and model governance as discretionary enhancements rather than mandatory infrastructure for AI use, particularly in SMEs without dedicated evaluation teams.
- Proprietary and Non-Transparent: The operational set of tools, frameworks, services, and integrations (tooling) often depends on individual vendor architectures, closed-source guardrails, proprietary logging mechanisms, and inaccessible safety layers. This prevents cross-model auditing, independent validation, or consistent system evaluation.
- Untraceable: Many current AI system deployments lack standardized event IDs, reconstructable decision trees, persistent logs of AI reasoning, and transparent information paths. Without internal traceability, post-incident investigations and compliance audits become impossible post hoc.

Internal governance remains uneven, opaque, and largely dependent on vendor configurations, leaving organizations unable to ensure or demonstrate system integrity.

Summary: A Fragmented Governance Landscape

Across all three domains, current governance systems fail to provide the real-time oversight, cross-layer alignment, and traceability required for modern AI-systems. The landscape is disjointed:

- Collective Governance is slow and reactive.
- Relational Governance is inconsistent and not auditable.
- Internal Governance is optional and untraceable.

This inherent fragmentation and core systemic failures create gaps that prevent organizations from safely scaling AI use or implementing demonstratable compliance with emerging regulatory frameworks. The Integrated Governance Stack (IGS) is designed to close these gaps.

The Integrated Governance Stack (IGS)

The IGS is designed to resolve the fragmentation, inconsistency, and operational gaps that currently define AI oversight across industries. Instead of treating governance as a

collection of isolated channels, IGS established a unified, multi-layer architecture that integrates collective, relational, and internal governance into a single coherent system. Its purpose is threefold:

1. To provide organizations with a structured, auditable, and scalable governance framework.
2. To align human-AI collaboration with regulatory expectations and safety best practices, and
3. To make effective AI oversight accessible to organizations without dedicated AI governance and ethical compliance teams.

IGS provides an integrated model that connects three essential domains of AI oversight: collective, relational, and internal governance. By linking these domains through structured feedback loops, IGS transforms governance from a set of disconnected activities into a coordinated, end-to-end oversight framework.

As AI systems increasingly participate in the operational decision-making process of businesses, it becomes imperative to create standardized methods to evaluate, document, or audit human-AI interactions. IGS introduces clear oversight roles, standardized evaluation pathways, reconstructable decision-trees, and mechanisms for multi-agent auditing. This ensures that human judgment, AI output, and agentic behavior can be reviewed, reproduced, and validated against regulatory and ethical criteria. IGS clarifies boundaries, assigns accountability, and enables consistent and repeatable human-in-the-loop processes.

IGS embed governance into technical substrates of AI systems by promoting persistent data logging, event classifiers, structured metadata capture, transparent decision-path reconstruction, and accountable-system configuration practices. The internal structure enables organizations to move beyond optional or unexposed vendor-side controls and establish traceability and integrity at the system level, equipping all sizes of enterprises and federal institutions the ability to build customizable AI ecosystems. IGS is explicitly designed to provide deployable, resource-efficient governance infrastructure that allows smaller organizations to evaluate AI systems responsibly, satisfy regulatory expectations, document and audit AI use, scale human-AI collaborations safely, and close gaps normally requiring specialized governance teams. It acts as the foundational governance overlay, allowing all businesses to implement high-quality oversight with modest staff and without relying on proprietary, siloed, or inconsistent processes.

IGS supports alignment with major regulatory frameworks including:

- EU AI Act's risk tiers and oversight requirements

- NIST's AI Risk Management Framework (RMF)
- ISO/IEC 42001 AI management system controls
- OECD's AI Principles of transparency, accountability, and fairness
- Sectoral obligations such as in financial, healthcare, and cybersecurity
- State-level requirements such as Texas' TRAIGA

While IGS does not replace regulatory compliance to these frameworks, it provides organizational structure and auditability needed to implement them cohesively and effectively.

As AI systems evolve toward greater complexity, autonomy, multi-agent orchestration, and continuous learning, organizations require governance models that can adapt dynamically. IGS establishes the governance on which future safety mechanisms, watchdog systems, and advanced oversight can operate.