The Illusion of Control: Why AI Is Outpacing Enterprise Security Models & What to Do About It

Authors

Poonacha Kongetira
CEO, Classie AI Inc
(fmr) SambaNova, Google, Nvidia

Roland Cloutier
CEO, The Business Protection Group.
(fmr) Global CSO Tik Tok & ByteDance,ADP, EMC

Executive Summary

Most enterprises believe they have control over AI. In reality, they are operating with limited visibility, fragmented policy enforcement, and expanding exposure. Artificial Intelligence is scaling faster than any prior enterprise technology shift. Organizations are already deploying GenAI broadly and building internal Agentic platforms, yet many lack a coherent strategy to manage risk, enforce policy, and operationalize control. This paper outlines an outcome-based model for security, risk, and technology leaders to enable business competitiveness through safe AI adoption.

> If you cannot see how AI is being used, you cannot operationalize it

The Reality and Risk of Enterprise AI Adoption

AI adoption is moving along an accelerated, evolving S-curve, driven by the convergence of accessible models and AI reinforced development that have dismantled traditional barriers to entry. Decentralized individual usage, 90% of which is unsanctioned, has driven gains in productivity. Boards & executives have taken notice, and are mandating AI adoption. Consequently, AI is being woven into enterprise workflows, decision-making, and customer interactions at an unprecedented pace.

In this high-pressure environment where innovation accelerates, but visibility, governance, and control lag behind, we see organizations defaulting into three common operating patterns:

- **Unfettered Adoption:** Quick adoption without structured control deliver short term gains, but introduces unmanaged and invisible risk
- **Reactive Management:** Addressing issues as they emerge results in fragmented, unscalable solutions that increase medium term operational complexity
- **Strategic Procrastination:** Delaying meaningful engagement while waiting for stable standards to emerge delays learning while AI continues to evolve rapidly

> Waiting to build an AI governance strategy is a competitive choice

The Three Mandates of the Modern Security Executive

For CISOs, CSOs, CIOs, and CTOs, AI introduces a set of overlapping responsibilities that must be addressed simultaneously:

- **Enable Business competitiveness:** The goal is to move from "No" to "Yes and here is how" on AI adoption, and to turn governance into a catalyst for innovation.
- **Defend the Enterprise:** Addressing issues as they emerge results in fragmented, unscalable solutions that increase medium term operational complexity
- **Strategic Procrastination:** Protect against a dual-threat landscape, AI-driven attacks on traditional infrastructure and entirely new categories of attacks targeting AI applications (such as prompt injection or data poisoning).

Balancing these mandates requires a shift from traditional control models to more adaptive, system-oriented approaches. To understand why, we must examine why AI is a fundamental departure from previous technologies.

Why AI is fundamentally different

Modern AI applications are built around trained neural networks like LLMs that differ substantially from traditional software:

- **The Black Box issue:** A trained neural network is a collection of billions of weights, numbers whose intent cannot be interpreted by 'reading weights' like a human could read code. And a large model cannot be pre-tested exhaustively.
- **The Variability issue:** Traditional code produces the same result each time. Models produce probabilistic results based on statistical likelihoods. Therefore models can produce variable responses to similar queries, and can be extremely suggestible to variations in prompting! Both can result in unsafe behavior from 'safe' AI applications, whether from prompt hacking or accidental misuse.
- **The Permissioning issue:** Old software workflows relied on human designed, point-to-point, deterministic interactions, with explicit permissions. Agentic workflows are different; an AI orchestrator spontaneously designs a path to solve a query, consuming and conveying data across systems without a pre-defined permissioning map. Because less constrained models often deliver superior results, the business pressure toward autonomy will only increase. This authorization hazard is a problem.
- **The Speed Issue:** AI is 100x faster than humans, is untiring, and can generate beneficial or catastrophic results that human oversight can never match for speed.

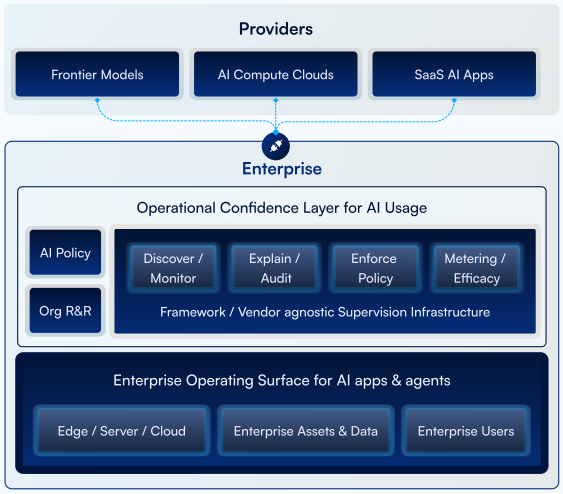
These attributes of AI require a **real-time supervisory function** to 'bless' or 'alert' actions taken by AI. In this AI resembles people. **To describe this, we use the term AI Supervision, as observability and perimeter security approaches used to manage traditional software are simply not enough.**

> **AI agents are like people — they need real-time supervision**

Further reading on threat models is strongly recommended: [1, 2]

The Operational Confidence Layer for AI

When adopting AI, the core business related operating agreements and policies of an enterprise must remain inviolable. To meet this requirement and the modern security mandate, organizations must establish an Operational Confidence Layer for AI—a reference AI policy architecture combined with operational capability that provides visibility, control, and enforcement across all AI usage. Crucially, the layer acts as a strategic decoupling point:



- It ensures portability, resilience, and consistency of enterprise policy enforcement across the full spectrum of AI manifestations on company infrastructure—sanctioned or unsanctioned, employee personal or business AI, internally developed or vendor supplied, self or vendor hosted—across diverse cloud and model providers.
- It ensures that the enterprise retains ultimate authority and custodianship over its data and actions, while providing strategic optionality across AI providers.

This layer does not seek to improve individual application performance which remains the preserve of development teams and vendors.

> **You buy the AI, but you must control your Operational Confidence Layer**

Core Supervisory Capabilities of the Confidence Layer

The Operational Confidence Layer is not a single view but a coordinated set of supervisory capabilities that create a control plane for the organization.

The bedrock capability is the systematic acquisition and compositing of contextually relevant data. Compositing combines different elements of interaction between users, AI systems, tools and data into a single cohesive map of the enterprise AI landscape. This enables real-time analysis and policy enforcement at "inference time," driving the following outcomes:

- **Discovery (The 3 C's):** Provides detection of sanctioned and "shadow" AI, continuously mapping the *Connectivity, Configuration, and Credentials* of every AI system, rolling up into a live Enterprise AI Posture Map that identifies vulnerabilities and exposure.
- **Explainability & Auditability:** Traces AI-driven outputs and actions while surfacing insights around intent, identity, chain of custody, and data lineage. It provides the forensic trail necessary for compliance, trust, and operational accountability.
- **Operational Security:** Utilizes Policy-as-Code integrating with existing infrastructure to enforce rules consistently. ML automation allows policies to be updated and deployed at machine speed, incorporating "human-in-the-loop" governance where critical oversight is required. It establishes trustworthy AI pipelines that integrate security controls throughout the lifecycle, ensuring code, data and model integrity from end to end.
- **AI Efficacy Metering:** This layer has all the information at hand to model AI operating costs and drive AI efficacy by comparing cost and outcomes between AI vendors. It must drive operational cost competitiveness to improve the bottom line.
- **Agile Durability:** Supervision capabilities must be durable. They must evolve as fast as AI application capabilities and AI attack vectors do while operating seamlessly across cloud platforms and endpoints, adaptable to different frameworks and vendors, preventing fragmentation and tool sprawl while maintaining a single source of truth.

Putting it all together

AI Supervisory capabilities, delivered frictionlessly and in real-time, are combined with Governance Policies to create an enterprise owned Operational Confidence Layer to harness AI and drive business competitiveness.

A Confidence Layer built on AI Supervision is Business Value

What Good Looks Like

In organizations that are effectively managing AI adoption, several patterns emerge. They:

- Have clear visibility into AI usage including both approved and unapproved tools.
- Enforce policy consistently, regardless of models, apps and deployments.
- Understand the operational and financial impact of AI usage and adjust seamlessly.
- Establish roles, processes, and technologies to handle AI risk without slowing innovation.

They achieve success by managing AI risk in a structured, transparent, and operationally enabled way.

What Leaders should do now

Recognize and communicate the strategic importance of the confidence layer as competitive differentiation for your company. Get executive buy-in, form a cross-functional task force, **start!**

01 **Begin with AI discovery:** Know what's running. Without the foundation of AI discovery and acquisition of contextually rich data, other efforts will be ineffective.

02 **Codify AI Policy:** Define 'north star' AI policies that bridge existing enterprise protections with the unique, probabilistic risks introduced by LLMs.

03 **Select some partners:** Free up budget and bandwidth to work with partners that can help build a confidence layer that you can own. Select partners for their durable vision and speed of execution on your issues. Waiting is the enemy.

04 **Deploy the Confidence Layer:** Get to a central point of control, ensuring supervision is decoupled from app, model and compute providers to give your business optionality.

05 **Modernize your operational workforce:** Recognize that AI requires new oversight roles, and train your teams to manage agentic workflows and automated ML systems.

This is an existential journey on an unknowable road.

Get started now!

References

[1] Mitre Atlas: Threat models for attacks against AI Systems
[2] ZeroDayClock: The collapse in time between CVE disclosure and confirmed exploitation