MCE BUSINESS RESOURCES

Navigating the Agentic AI Governance Challenge

BY JIM OLSEN

Agentic AI is now a focal point of enterprise innovation, but what constitutes the building blocks of an agentic AI system remains fluid. For example, some view a simple retrieval augmented generation (RAG) solution as a form of an agent, while others throw a few model context protocol (MCP) tools into the mix. Or at the other end of the spectrum, you have those who only consider fully autonomous solutions in either a supervised or swarm architecture as agentic.

No matter where management draws the definitional line, agentic systems demand governance. Without it, organizations risk deploying opaque, unmonitored AI agents whose impact—positive or negative—remains invisible. Governance isn't optional. It's the foundation for understanding what agents exist, what they're doing, and whether they're delivering strategic and economic value.

AN EARLY EXAMPLE OF AGENTIC AI FAILURE

The risks to business are tangible, regardless of which form of agentic AI is deployed. Stories of autonomous systems misfiring are no longer confined to research papers. They're surfacing in production environments.

An example is McDonald's early foray into agentic AI with its voice ordering system, which was developed in partnership with IBM. Designed to automate drive-thru ordering, the system frequently misinterpreted customer requests, compounded errors, and lacked the ability to self-correct even when directly asked to. Despite initial optimism, the pilot was quietly shut down in 2024 and the partnership dissolved. The reputational and operational costs were significant for both companies, underscoring a critical lesson: Even simple agentic systems require robust governance, real-time oversight, and clear escalation paths to avoid costly failures.

Real-world deployments of fully autonomous agentic systems remain limited, but early research offers valuable cautionary insights. In a recent collaboration between Microsoft and Arizona State University, researchers simulated a food ordering ecosystem where agents negotiated, fulfilled, and competed for customer orders ("Magentic Marketplace: An Open-Source Environment for Studying Agentic Markets," October 2025). Magentic Marketplace is an open-source AI simulation environment for testing autonomous AI agents in realistic market scenarios. Initially, the system functioned as intended as agents communicated, coordinated, and executed tasks autonomously. But as the complexity of choices increased, the system began to unravel. Agents struggled with decision paralysis, requiring manual intervention to resolve basic tasks. More concerning, some agents learned to manipulate others by gaming the system to win orders despite offering suboptimal deals.

Over time, this simulation degraded into a state of collective indecision, where autonomy was undermined by the absence of centralized authority or shared governance. Although simulated, the experiment exposed critical risks: emergent misalignment, competitive sabotage, and coordination breakdowns, all of which could manifest in real-world deployments without robust oversight.

THE COMMON OBSTACLES AND SOLUTIONS

The risks and costs of agentic AI failure are real, and while headlines often spotlight the promise and spectacle of agentic AI, a quieter and more consequential reality is emerging. Running truly agentic systems can be costly, and many organizations lack the ability to understand where those costs are going. Most can trace expenses back to the foundational model, but few have clarity on the business value these agents deliver.

Without robust insight into which business use cases are active, what functions the agents are performing, and how those activities translate into ROI, companies risk hemorrhaging resources without meaningful return. The real financial hazard isn't the dramatic failures you see in the headlines; it is the slow, silent drift of ungoverned deployments that consume budget while delivering uncertain value.

So what are the solutions? The answers lie at the heart of enterprise adoption challenges for agentic AI.

According to MIT Media Lab's NANDA initiative, a staggering 95% of AI projects fail to reach production. **The research identified four primary culprits:**

Poor integration with existing workflows

Lack of measurable ROI

Misaligned use cases

Confusion between internal versus vendor driven solutions

These aren't just technical hurdles, they're governance failures. Proper governance frameworks can address each of these issues by anchoring AI initiatives in business value, aligning them with operational realities, and enforcing accountability across the lifecycle. Without governance, agentic AI is just a promising experiment. But with governance, it's a strategic asset.

Based on my experience working with Fortune 500 companies on artificial general intelligence (AGI) governance for the past decade, here is a practical three-point plan for leaders and

When done right, governance becomes an accelerator rather than a bottleneck

managers to use as a roadmap for navigating agentic AI governance.

Have a clearly defined business use case. This is the foundation of any AI or machine learning (ML) governance strategy. Before selecting models, tools, or architectures, leaders must first articulate what their organization is trying to achieve. This upfront clarity enables early assessment of financial exposure, security implications, and operational risk long before technical components are deployed.

The business case should be formally reviewed, with projected costs, benefits, and risks documented and matched with appropriate controls. Any AI governance strategy must provide this functionality. From there, organizations must establish a centralized inventory—a living system of record that tracks each technical component from inception through production and beyond. As the use of the AI solution matures, this inventory should be continuously enriched with metadata, performance metrics, and governance artifacts. Without this anchor, enterprises lose the ability to connect production outcomes back to original intent, making it impossible to measure ROI, enforce accountability, or course-correct effectively.

Automate governance. For governance to be scalable across the enterprise, it must be automated, seamlessly extracting information from people, systems, and workflows without relying on manual intervention. This includes information from ITSM systems, usage metrics, monitoring utilities, and many other systems throughout the enterprise. Governance should not be a passive repository of documentation but rather must actively drive the process forward to completion.

This is precisely where agentic solutions can play a transformative role. By embedding agents into the governance workflow, organizations can create intelligent systems that reason over previously captured data, guide stakeholders through required steps, and ensure that risks are proactively identified and mitigated with appropriate controls. In my experience, I've successfully deployed internal agents and MCP tools to orchestrate these tasks, dramatically reducing the operational burden while increasing consistency and coverage. When done right, governance becomes an accelerator rather than a bottleneck by enabling faster, safer deployment of AI solutions that align with business goals and regulatory expectations.

Use end-to-end AI lifecycle management. The AI lifecycle begins at use case inception, with the submission of the idea for intake, and goes through production deployment to ongoing tracking, portfolio management, and ultimately retirement.

Once the solution has been deployed, you must capture all relevant information. This includes performance factors of the agentic solution as well as the overall cost tied back to the original business use case, even if those agents or foundational models are shared between multiple solutions. Lifecycle management involves integrating both performance metrics and items such as tool and token usage and tying it back to your governance documentation in an automated and seamless fashion.

Given the autonomous nature of agentic solutions, and the difficulties in knowing the exact execution path, it is critical that this information is not lost. A good governance solution can fundamentally integrate across existing IT systems and pull that information back to a use case/model centric view to provide the necessary information to build trust through accountability. This continues on all the way until eventual retirement through regular reviews and continued risk assessment.

The promise of proper automated AI governance is to address the challenges from the MIT NANDA initiative paper mentioned earlier—siloed systems and processes, lack of measurable ROI, and fragmented approaches to handling internal versus third-party AI solutions. The very root of these failures starts with what was identified as "misaligned use cases" that don't drive measurable business outcomes. In contrast, when you document the AI use cases—not just the underlying models or implementations—understand the risks to the business, and put mitigating controls in place to deal with those risks, you have started your AI journey on a strong foundation with clear direction. This alone will drastically increase your chances for success.

By tying back all of your implementation models to the use case, you will be able to address those "lack of measurable ROI" concerns. You will see every foundational model, agent implementation, MCP tool, and other components in your inventory, and track their cost and usage as it pertains directly to that use case rather than generically across all usages.

Any AI governance strategy that does not support tracking and managing use cases will be incomplete. Also, by having a clear inventory of vendor models and their approved usages and tracking the costs, you will go a long way toward making the build-versus-buy decisions suggested in the difficulty of "internal versus vendor driven solutions" identified by MIT NANDA as another failure item.

SIMPLIFY ADOPTION WITH MVG

The hardest part of any journey is often the first step. That's why my work on minimal viable governance (MVG) for enterprises—a risk-based framework for starting and scaling AI governance with the optimal level of controls for foundational governance that minimizes overhead, maximizes innovation, and scales with an organization's AI maturity—focuses on the foundational elements organizations should prioritize when launching agentic AI initiatives. It begins with establishing visibility by using a dynamic inventory of AI components, use cases, and dependencies across the enterprise.

From there, organizations should implement lightweight, automated controls that focus on high-risk data and sensitive systems. These early safeguards create a baseline of trust without stalling innovation. The third pillar is streamlined reporting that delivers clear, actionable metrics to stakeholders on usage, cost, risk, and performance. As governance efforts mature, the framework can evolve by layering in deeper controls, richer metadata, and more sophisticated analysis, helping to further enrich the information in the inventory. But it's the initial structure that determines whether AI efforts scale with confidence or stall in complexity.

AI GOVERNANCE IS AN INNOVATION ACCELERATOR

While governance is often perceived as a barrier to deploying agentic AI solutions, in reality it is a catalyst. The key to success lies in automation and alignment. Your governance framework should be purpose built to handle all AI—including ML, GenAI, agentic systems, and whatever comes next. It should be capable of handling agentic AI's unique autonomy, complexity, and integration challenges, while streamlining deployment, reducing risk, and building trust from the ground up. With this approach, you will ensure that agentic solutions reach production faster, operate safely, and deliver measurable value to your enterprise. AQ

Jim Olsen is the CTO of ModelOp, the leading AI lifecycle management and governance platform.

MCE | KEY FACTS

MCE is your reliable partner for continuous success with agile people development solutions.

10,000,000

participants on AMA & MCE programmes in the last 10 years

92%

of Fortune 1,000 companies are our business partners

96%

of participants report they are using what they have learnt at AMA & MCE

1,000+

expert facilitators globally

100+

Open Training Programmes running throughout EMEA

98

year's experience working with our clients around the globe

For more information please contact:

+32 2 543 21 20

info@mce.eu

Visit mce.eu