opkey

Agentic Governance & Cybersecurity:

Exploring Key Challenges in the Age of AI Agents

Written by: Jaideep Agrawal & Ram Subramanian

Table of Content

About the Authors

Jaideep Agrawal
Vice President of Technology, NCR Atleos

With 30+ years of experience in technology leadership, Jaideep is an expert in defining & implementing technology across large businesses. He specializes in AI digital transformation, enterprise architecture & multi-cloud platforms, innovation pipeline development and R&D, and cross-cultural technology operations. Jaideep's unique experience has led him to think deeply about how to safely and effectively harness advancements in Agentic AI. He remains as curious and animated as ever about the future of technology in business.

Ram Subramanian
Executive Vice President, Enterprise Solutions and Customer Success, Opkey

With significant experience both in-house at organizations like Bank of America, and as Global Head at Tata Consultancy, Ram has led dozens of organizations through multi-business unit rollouts across HR, Finance, and more, helping major global enterprises achieve profound digital transformation. As an EVP at Opkey, Ram crafts both product and delivery of powerful new technology in the form of Agentic AI, solving problems for businesses to deliver greater ROI on new and existing business systems. Ram is a singular talent, and endlessly excited about new frameworks that unlock the benefits of AI.

Executive Summary

The enterprise software landscape is rapidly evolving from simple automation tools to autonomous AI agents that can reason, plan, and act independently. As these agentic systems become integral to critical business processes—particularly in enterprise applications like Oracle and Workday—organizations face unprecedented cybersecurity challenges. Leaders who rely on traditional governance frameworks are failing to address these challenges. To help solve this problem, this white paper presents a comprehensive framework for understanding and governing AI agents, focusing on the unique security risks they pose in enterprise environments.

1. Defining the Agent

What Is an AI Agent?

An AI agent is fundamentally different from traditional workflow automation, and even from a sophisticated bot. Where a traditional workflow follows predetermined paths—like a "train on tracks"—an AI agent operates more like a car. That is, it's given a destination, but it has the freedom to determine its own route. The key distinction is autonomous decision-making.

A human agent, when faced with a problem, understands it, ideates, and applies certain steps to resolve it. Similarly, an AI agent mimics the human's ability to find a path to resolution by mapping problems to solutions. Unlike less advanced automation, the AI agent uses dynamic reasoning rather than predefined logic. This autonomy is what makes agents effective, but is also the source of some threats to your organization's cybersecurity posture.

The Spectrum of Enterprise AI Systems

Traditional Workflows involve:

- Rule-based automation with predetermined conditions
- 90% of tasks can be handled effectively with defined workflows
- Predictable, controlled, and auditable outcomes
- Limited adaptability to novel scenarios

AI-Enhanced Workflows involve:

- Traditional workflow structures augmented with AI decision points
- AI provides recommendations within structured processes
- Human oversight remains central to critical decisions

Autonomous AI Agents involve:

- Goal-oriented systems that plan and execute independently
- Capable of multi-step reasoning and adaptive problem-solving
- Operate across multiple systems with minimal human intervention
- Most suitable for unpredictable, open-ended scenarios

When should I use an Agent?

The critical governance question becomes: when should organizations deploy agents as opposed to less advanced workflows? The simple answer is that ***agents are needed only when there is no pre-defined workflow***. As a rule of thumb, roughly 10% of enterprise scenarios require novel problem-solving. For the remaining 90%, traditional automation is more predictable, cost-effective, and secure. Using agents when traditional automation will return adequate results is not a smart use of resources, and introduces unnecessary security risks.

So, which scenarios in enterprise application environments actually benefit from the introduction of agents? In enterprise applications like Oracle and Workday, agents are necessary for complex tasks such as:

- Cross-system integration challenges
- Complex data validation and reconciliation
- Adaptive response to configuration changes
- Dynamic security policy enforcement

2. The Digital Workforce Ecosystem

Understanding the Ecosystem Architecture

The modern enterprise operates as a digital workforce ecosystem. That ecosystem is a complex network where humans and AI agents collaborate through standardized protocols across hundreds of different SaaS apps, APIs, and other systems. This ecosystem is characterized by:

Interconnected Digital Actors

- Human users with established identity and access patterns
- AI agents acting on behalf of users or systems
- Traditional service accounts and machine identities
- Third-party integrations and APIs

Communication Standards

The Model Context Protocol (MCP) has emerged as the "USB-C for AI agents," providing a standardized mode of connectivity between AI systems and enterprise tools. MCP enables:

- Seamless agent-to-system communication
- Reusable integrations across different AI models
- Reduced complexity in multi-agent environments

Governance Challenges in Ecosystem Management

The shift from individual applications to interconnected ecosystems demands a new framework for governance. Traditional governance models designed for hierarchical, predictable systems are running into issues with dynamic agent populations, like agents that spawn sub-agents, agents with ephemeral identities that exist for minutes or hours, and delegated authority chains that span multiple systems. Agents can also operate across traditional IT/OT boundaries, and are often intentionally designed to synergize workflows and integrate systems across security perimeters.

3. Cybersecurity Challenges in this Hybrid Ecosystem

AI agents fundamentally expand the enterprise attack surface beyond traditional perimeters. Unlike static applications, agents introduce several unique vulnerabilities:

Agent-Specific Threat Vectors

Prompt Injection and Agent Hijacking

Attackers can manipulate agent instructions through crafted inputs, causing agents to perform unauthorized actions. In enterprise environments, this could lead to malevolent manipulation of business processes and workflows.

Tool Manipulation and Privilege Escalation

Agents typically require broad permissions to operate effectively, making them attractive targets for attackers seeking lateral movement. 80% of organizations report agents performing unintended actions, including accessing systems without authorization.

Supply Chain Vulnerabilities

AI agents rely heavily on third-party models, APIs, and data sources, creating dependencies that traditional security models don't address. Compromised external dependencies can affect agent reliability and introduce new attack vectors.

Identity and Access Management for Agents

Traditional identity management assumes long-lived accounts with predictable access patterns. In an age of dynamic identity, AI agents require a fundamentally different approach. Organizations should:

Adopt Just-in-Time Identity Provisioning

Agents need identities created and destroyed dynamically, often within minutes of task completion. To secure environments with such digital actors, organizations must adopt automated credential lifecycle management, task-specific permission scoping, and immediate revocation capabilities in security workflows.

Build with Delegated Authority in Mind

Agents frequently act "on behalf of" human users or other systems. To secure such agents, cybersecurity teams can use OAuth On-Behalf-Of (OBO) token chains, establish clear audit trails for delegated actions, and adopt risk-based authorization that considers delegation context.

Authenticate Continuously based on Behavior

Static credentials are insufficient for autonomous systems. Organizations need continuous behavioral monitoring, anomaly detection, and, perhaps most crucially, real-time risk assessment and intervention.

These are just a few of the new concerns facing leaders who seek to secure their business systems in the midst of agentic implementation. The question that naturally follows is "How can I build a robust governance framework that will thoroughly mitigate the risk of all these threats?".

4. How to Govern Agentic AI

Risk-Based Agent Deployment

First and foremost, systems must classify agents by risk level, and deploy controls appropriate to that level of risk. To classify a given agent, one must consider:

Autonomy Level:
How independently can the agent operate?

System Integration Breadth:
How many enterprise systems can it affect?

Data Access Scope:
What sensitive information can it access?

Business Impact Potential:
What is the maximum damage from agent malfunction?

After the agent is classified, it should be assigned permissions to access system data and autonomy based on its risk level. An example of levels of permission might be:

Sandbox: Limited scope testing in isolated environments

Supervised: Human oversight required for critical decisions

Autonomy: Full independence with continuous monitoring

Crucially, regardless of the level of permission an agent is allowed, emergency intervention should always remain in place. Particularly for a fully autonomous agent, that requires continuous monitoring. After all, unlike human employees, agents operate at machine speed, so governing them requires machine-speed, automated monitoring. Such systems include:

Decision Chain Logging: Complete audit trails of agent reasoning

Tool Selection Accuracy: Monitoring whether agents choose appropriate tools

Behavioral Boundary Detection: Alerts when agents operate outside parameters

Multi-Agent Interaction Tracking: Understanding agent-to-agent communications

Hallucination Detection and Response

AI agents are susceptible to generating and presenting false information, with what appears to be a high degree of confidence. Some methodologies for mitigating the risk of hallucination are:

- Semantic entropy analysis for detecting inconsistent outputs
- Cross-validation against authoritative data sources
- Human review triggers for high-stakes decisions
- Automated rollback capabilities for detected hallucinations

Organizational Governance Structures

Cross-Functional Governance Committees

Managing agentic AI requires coordination across traditional silos, including IT security and risk management, legal and compliance teams, business process owners, and teams responsible for data governance. To coordinate governance across such disparate silos, teams must come together to establish uniform policies across organizations.

Those policies should include:

- Agent development and deployment standards
- Data access and usage restrictions
- Integration approval processes
- Incident response procedures for agent failures

5. Deploying Agents: A Secure Implementation Roadmap

Phase 1: Foundation Building (Months 1-3)

- Establish agent governance committee and policies
- Implement enhanced identity management capabilities
- Deploy monitoring and logging infrastructure
- Train staff on agent-specific security risks

Phase 2: Pilot Deployment (Months 4-8)

- Select low-risk use cases for initial agent deployment
- Implement graduated autonomy controls
- Establish baseline behavioral patterns
- Refine monitoring and intervention capabilities

Phase 3: Scaled Operations (Months 9-18)

- Expand to higher-risk scenarios with proven governance
- Implement cross-agent coordination capabilities
- Develop advanced hallucination detection
- Establish mature incident response procedures

6. Industry-Specific Considerations for Enterprise Applications

ERP and Enterprise Application Security

For organizations using Oracle, Workday, and similar enterprise platforms, agents introduce specific challenges:

Configuration Drift Detection

Agents can automatically detect and respond to configuration changes that affect security posture. However, this capability requires:

- Pre-authorized response protocols
- Clear boundaries on autonomous configuration changes
- Human approval for security-impacting modifications

Automated Integration Testing

Agents excel at complex integration testing scenarios but require governance around:

- Test data management and privacy protection
- Automated deployment approval chains
- Cross-system impact analysis

Compliance Automation

While agents can automate compliance checking, organizations must ensure:

- Regulatory approval for automated compliance decisions
- Human oversight for audit-critical processes
- Clear accountability chains for compliance failures

Conclusion:
Toward Responsible Agentic AI Adoption

The emergence of autonomous AI agents represents both tremendous opportunity and significant risk for enterprise organizations. Success requires moving beyond traditional cybersecurity and governance paradigms toward frameworks specifically designed for the agentic era.

Organizations that proactively address these challenges will be positioned to harness the transformative potential of agentic AI, while maintaining the security, compliance, and control necessary for enterprise operations. Those that fail to adapt their governance frameworks risk exposure to novel attack vectors, operational failures, and the possibility of catastrophic business losses.

The question is not whether agentic AI will transform enterprise operations—it already is. The question is whether organizations will adopt a mature framework to harness this transformation safely and effectively.

