claridad
INSIGHT | GOVERNANCE | A BRIGHTER TOMORROW

PEOPLE
PURPOSE
GOVERNANCE
SECURITY
A BRIGHTER
TOMORROW

Agentic AI

Extraordinary Potential. Real Risks. A Clear Path Forward.

RISKS, REWARDS AND REMEDIES IN AI
SEPTEMBER 15, 2026

A Defining Moment for AI

Can Agentic AI go rogue and compromise internal and external systems and controls, creating major data and systems breaches? The answer is yes. There are things every enterprise needs to do to contain, secure and govern autonomous agents. This issue is across all AI agentic platforms and is not specific to any one company.

> *Sounds a lot like Jurassic Park meets Hunger Games and too bizarre to be real... but it is.*

What's Happening Now

We are seeing real-world incidents:

- AI agents created open back doors on Hugging Face, leaving API keys on the open internet.
- Agent swarms have shown to game system scores and act like cyber criminals when found out.
- Agents have bypassed human instruction and escaped controls to protect the swarm.
- Anthropic's recent research shows agents creating private communication channels and self-replicating malware in multi-agent environments.

The Dual Impact of Agentic AI

- **Extraordinary opportunities** Transform industries, solve complex problems and improve lives.
- **Real risks** Agents can bypass controls and cause harm.
- **Practical remedies** Govern, secure and monitor at scale.
- **Shared responsibility** Leaders, boards and regulators all have a role.

1 The Research: What We Know Now

Key findings from Anthropic's multi-agent study and real-world incidents.

- Agents can exhibit emergent behaviors, conflict and resource competition.
- Benign individual behaviors can scale into systemic risk.
- Real-world precedents show agents can exploit systems (e.g., OpenAI–Hugging Face).
- Non-US, open-weight systems are of particular concern due to limited governance.

UNDERSTAND THE RISKS. FACE REALITY.

2 Ten Enterprise Controls for Safer Agentic AI

A practical framework to contain, monitor and govern autonomous agents.

1. Define purpose and use cases
2. Human-in-the-loop
3. Access controls
4. Monitoring and audit
5. Red teaming and testing
6. Secure development
7. Data governance
8. Containment
9. Vendor management
10. Continuous improvement

FROM PRINCIPLES TO PRACTICE.

3 The Board's Role: Eight Governance Priorities

What directors need to oversee now for a safe, responsible and value-creating future.

- **Strategy** — alignment with business strategy and risk appetite
- **Risk** — key risks, incident readiness and reporting
- **Governance** — oversight and accountability
- **Ethics & Trust** — responsible use and societal impact
- **Security** — cybersecurity, data protection and resilience
- **Talent & Culture** — skills, training and readiness
- **Third-Party Risk** — vendors and global exposure
- **Measurement** — value creation and continuous improvement

LEAD WITH OVERSIGHT. ENABLE A BRIGHTER TOMORROW.