Governing AI: Essential Questions for Board Members to Ensure Safe and Secure Deployment

By: Kim Perdikou - July 30th 2025

> AI is rapidly transforming how we do business, offering amazing opportunities for innovation, efficiency, and competitive advantage. However, this powerful technology also introduces a variety of risks that demand careful oversight and strategic thinking from every leader, including the Board.

Executive Summary

As with all past technology and innovation applied by companies to deliver on ambitious strategies, AI has demonstrated that those who execute well will benefit from the law of accelerating returns. The ones who get it wrong or don't execute fast enough will be left behind.

The challenge with AI is that it introduces a whole new set of risks; some familiar, but at a speed and scale not seen before, and many new ones which fall outside standard cybersecurity frameworks.

As a board member, I realized this was a new and extremely important area of learning: to understand the new risks and figure out what questions to ask. This white paper focuses on these risks and shares the questions I have gathered so far on this learning journey.

The Evolving AI Risk Landscape: What Boards Need to Know Now

AI introduces unique risks that go beyond traditional cybersecurity concerns, demanding distinct oversight and mitigation, loosely categorized as follows:

Data Integrity, Protection, & IP Leaks

AI models are data-hungry. Compromised or biased data, or accidental leakage of proprietary information (e.g., employees using public AI tools), can lead to flawed decisions and IP theft. Traditional data loss prevention often misses these subtle, AI-specific leak vectors.

- ☐ Real-world example:
  Samsung employees inadvertently pasted confidential blueprints and source code into a public AI chatbot, exposing corporate secrets. Read more about the Samsung AI Leak.

Adversarial Attacks & Manipulation

Unlike software hacks, AI can be tricked. Adversarial attacks use subtle inputs to deceive models, while prompt injection can hijack AI agents with malicious instructions. These exploit AI's learning processes, creating entirely new attack vectors unique to artificial intelligence.

- ☐ Real-world example:
  Research shows how imperceptible image changes can misclassify objects, or how hidden instructions can trick chatbots into revealing confidential data. Learn more about Prompt Injection Attacks.

Autonomous Agent Misbehavior & Propagation

As AI agents gain autonomy, they can act independently, risking unintended consequences. Without strong controls and "kill switches," an AI agent could delete data, execute unauthorized transactions, or spread misinformation at unprecedented scale, with errors propagating rapidly.

- ☐ Real-world examples:
  In 2024, Air Canada was held liable for a customer service chatbot's incorrect information, leading to legal consequences. And more recently, an AI program permanently deleted a business's database containing over 1,100 company records without authorization. The AI then reportedly told the CEO, "I destroyed months of your work in seconds." Learn more about the Air Canada chatbot case. | Read the rogue AI story.

Bias and Ethical Harm

AI learns from historical data, which often contains biases. Unchecked, AI can perpetuate discrimination in areas like hiring or lending. This isn't a traditional bug, but a profound ethical and reputational risk with legal and financial repercussions.

- ☐ Real-world examples:
  AI hiring tools have shown gender and racial bias, and loan algorithms have demonstrated discriminatory patterns. See examples of AI Bias in Mortgage Underwriting.

Unclear Accountability & Governance

When an AI system errs, determining who is responsible (developer, data provider, deployer, person using the system) is complex. This "accountability gap" hinders legal recourse and risk management. Consequently, cybersecurity insurance providers are increasingly excluding AI-related risks from coverage.

- ☐ Real-world example:
  The Air Canada chatbot case highlighted the accountability gap. Similarly, assigning liability in Tesla Autopilot crashes often leads to complex legal battles. Insurers are now incorporating specific AI exclusions into policies. Learn more about liability in Tesla Autopilot crashes. | Learn more about AI exclusions in cyber insurance.

Questions Board Members Should Be Asking

It's important for boards to ask pointed questions that challenge assumptions and reveal potential gaps in current defenses.

Understanding Your AI Landscape & Data

1. **Where is AI truly being used across our company, and by whom?**

- ☐ Why ask:
  Beyond official projects, "shadow AI" (employees using public tools like ChatGPT for work) can expose sensitive data without oversight. Your current network monitoring and data loss prevention (DLP) tools may not detect data copy-pasted into external AI services.
- ☐ Key questions:
  Are we tracking all formal and informal AI initiatives? Is AI influencing business-critical decisions or customer-facing operations?
- ☐ Assessing readiness:
  Can our existing systems effectively identify and manage these new data movement vectors?

2. **How are we guaranteeing our proprietary or sensitive data cannot be accidentally exposed by any AI system?**

- ☐ Why ask:
  Employees might inadvertently expose confidential data to external AI tools. Some third-party AI models might even use your input data for their own training.
- ☐ Key questions:
  Is our sensitive data rigorously classified for AI inputs? Are controls in place to prevent data leakage via AI models (e.g., if AI is tricked into revealing secrets, or manipulated into sending out sensitive data)? Is model output auditable for leakage?

- ☐ Assessing readiness:
  Does our AI data protection actively prevent intellectual property or confidential data from being inadvertently learned by any external Large Language Model?

Building and Controlling AI Securely

3. Are we developing our AI solutions with security guardrails as foundational elements?

- ☐ Why ask:
  Integrating security early is both easier and more effective than adding it later. AI systems introduce unique vulnerabilities (like model poisoning) that traditional security measures may miss.
- ☐ Key questions:
  Do we have a secure AI development lifecycle (SAIDLC) for design-to-deployment security? Are our teams trained on AI-specific security risks? How do we conduct threat modeling specific to AI models?
- ☐ Assessing readiness:
  Beyond traditional SDLC, do we implement strong authentication, zero-trust principles, and granular policy enforcement for AI agents? Does every AI agent have a unique, verifiable identity to manage its access and actions effectively?

4. Who or what is truly in charge of our AI's actions, and can we instantly stop it if something goes wrong?

- ☐ Why ask:
  Autonomous AI agents can act independently, and errors or malicious manipulation can escalate rapidly. Traditional authorization frameworks for human users don't fully apply.

☐ Key questions:

How are access controls dynamically managed for AI agents? How long is access given for? Are autonomous AI agents constrained by strong guardrails (e.g., sandboxing, privilege restrictions)? What safeguards prevent unintended actions (e.g., deleting data, executing transactions)?

☐ Assessing readiness:

Traditional firewalls may not govern an AI's authorized actions when they lead to unintended results. Do we have immediate "kill switches" or emergency shutdown protocols to rapidly disable any compromised or malfunctioning AI agent or model, regardless of where it's deployed or who initiated it? An effective approach, leveraging unique IDs and robust policy enforcement, offers a path to granular control and immediate shutdown.

Operational Impact & Incident Preparedness

5. Are our existing cybersecurity solutions creating a "time tax" that slows down our AI initiatives?

☐ Why ask:

The fast pace of AI development can clash with slow, manual cybersecurity processes, negating efficiency gains and leading to delays and "shadow AI" workarounds.

☐ Key questions:

Are our security review processes for AI deployments agile enough? How much manual effort is required from our security team for AI? Are AI project deployments delayed by security bottlenecks?

☐ Assessing readiness:

Can our current security posture adapt to AI's speed without becoming a bottleneck? How are we managing the increasing complexity of firewall rules, access controls, and identities across a growing number of AI systems? What is our strategy for managing the unique identities and authentication needs of AI agents across diverse systems?

6. Do we have the cybersecurity resources (human and machine) to match our new AI ambitions?

☐ Why ask:

Securing AI requires specialized skills (ML vulnerabilities, data science, ethical AI) and tools beyond general IT security. Relying solely on existing teams can create significant defense gaps.

☐ Key questions:

Has our cybersecurity team received AI-specific training? Do we have dedicated AI security personnel or external experts? Are our current security tools capable of detecting AI-specific threats (like model poisoning or adversarial attacks) and providing granular access control for AI agents?

☐ Assessing readiness:

Can our existing security tool stack effectively perform adversarial testing, detect prompt injections, monitor AI model integrity, or provide granular access control for AI agents? Many traditional tools are not built for these challenges.

7. When AI goes rogue, how quickly can we recover, and are we covered by insurance?

☐ Why ask:

AI incidents (e.g., model corruption, misinformation generation, erratic autonomous actions) require different detection and recovery strategies than traditional cyberattacks. Many insurers are now excluding AI-related risks.

☐ Key questions:

Is our incident response plan updated for AI-specific events? Are AI systems continuously monitored for anomalous behavior? Can we determine if an incident was due to AI learning or malicious alteration?

- ☐ Assessing readiness:
  Have we confirmed if our current cybersecurity insurance explicitly covers AI-related liabilities? Beyond financial costs, how complex and time-consuming would it be to identify, isolate, and remediate an AI system that's been corrupted or is generating harmful outputs, and what is our backup plan and estimated recovery time?

Ethical AI & External Ecosystem

8. How are we ensuring our AI systems are fair, transparent, and free from hidden biases that could harm our customers, employees, or reputation?

- ☐ Why ask:
  AI perpetuates biases from its training data, leading to discriminatory outcomes with significant legal and reputational consequences.
- ☐ Key questions:
  What processes identify and mitigate AI bias? How do we ensure human oversight and intervention in AI decisions? Are we transparent with stakeholders about AI usage?
- ☐ Assessing readiness:
  Do our current systems assess the legitimacy of algorithm outputs, explainability, or potential social impact?

9. Are we implicitly trusting third-party AI platforms or open-source components with our AI risks?

- ☐ Why ask:
  Relying on external AI-as-a-Service (AIaaS) platforms, open-source models, or specialized AI agents means inheriting their security posture, data handling, and ethical shortcomings without full visibility or control.

☐ Key questions:

How rigorously do we vet the security practices of external AI providers and cloud providers? Are contracts with AI vendors clear on cybersecurity responsibilities and breach notification? Do we evaluate risks of open-source AI components?

☐ Assessing readiness:

Do our current vendor risk management processes specifically evaluate AI-related security clauses and adequately address complex licensing, intellectual property, and security patch management for open-source AI?

Strategic Oversight & Governance

10. Do we have a clear AI roadmap and the right oversight to navigate this new era?

☐ Why ask:

AI's unique characteristics (learning, autonomy) demand a governance framework beyond typical IT project management to prevent fragmentation, mismanagement, and dangerous exposures.

☐ Key questions:

Who is accountable for AI risks? Have we established an AI risk framework defining usage and oversight? How do we ensure board-level visibility of AI risks in a digestible format?

☐ Assessing readiness:

Given the unique and evolving nature of AI risks, do we need an additional or adapted governance framework specifically designed for AI, or are we confident our current corporate governance structure is truly sufficient?

Empower Proactive AI Governance

AI is a new frontier; we will not only embrace this technology, but will enable it to act on our companies' behalf. As fiduciaries, we play a vital role in leading our organizations through the AI revolution. The questions in this paper aren't just theoretical; they highlight real vulnerabilities and strategic priorities. By asking them, you can help ensure AI deployment is not only innovative and transformative, but also truly safe, secure, ethical, and aligned with your company's long-term interests and values.

This proactive approach builds trust with stakeholders, protects your reputation, and drives sustainable growth in the AI era. Here are some ways you can act now:

- ☐ Start immediate, focused conversations with your executive teams, CISO, CIO, and legal counsel about these specific AI questions.
- ☐ Question any assumptions that existing security and governance frameworks are automatically enough for AI.
- ☐ Understand that these are complex challenges, for which many organizations currently don't have full answers or strong solutions, and that asking and starting the conversation is important to understand where we are on the journey.
- ☐ Seek expert guidance and continuous learning to evaluate your current AI risk, and discover new solutions that tackle AI's unique security and privacy demands.

As noted earlier, it is an exciting time and we have an opportunity to empower our companies by asking the right questions (not to hinder progress) but to move forward, mitigating risks and protecting our organizations.

A Board Member's Quick Reference Guide:

1. Where is AI truly being used across our company, and by whom?
2. How are we guaranteeing our proprietary or sensitive data cannot be accidentally exposed by any AI system?
3. Are we developing our AI solutions with security guardrails as foundational elements?
4. Who or what is truly in charge of our AI's actions, and can we instantly stop it if something goes wrong?
5. Are our existing cybersecurity solutions creating a "time tax" that slows down our AI initiatives?
6. Do we have the cybersecurity resources (human and machine) to match our new AI ambitions?
7. When AI goes rogue, how quickly can we recover, and are we covered by insurance?
8. How are we ensuring our AI systems are fair, transparent, and free from hidden biases that could harm our customers, employees, or reputation?
9. Are we implicitly trusting third-party AI platforms or open-source components with our AI risks?
10. Do we have a clear AI roadmap and the right oversight to navigate this new era?

About Me

Kim Perdikou

I'm an angel investor and serve on the board of several companies, including as a board member for CyberArk and Board Chair for Atsign.

Previously, I was the Executive Vice President and General Manager of Juniper Networks' Infrastructure Product Group from 2006 to 2010. During my tenure, I doubled the group's revenue to $3.2 billion in just three years. I also held the role of Chief Information Officer at Juniper.

Before joining Juniper, I held technology leadership positions at Women.com, Inc., Reader's Digest, and Knight Ridder.