October 8, 2026

Cybersecurity Risk Oversight in the Age of AI

As advanced AI models — and autonomous AI "agents" — grow more capable, so do the associated corporate cybersecurity risks. At increasingly modest cost, AI systems can now identify and exploit latent software vulnerabilities, such as by planting malicious code or stealing data, at a speed and scale beyond the reach of human hackers or conventional automated tools. AI also creates new points of vulnerability: a corporation's own AI agents, if compromised, can become a gateway to sensitive data and to operational or payment systems. The risk of deepfakes (AI-generated voice or video impersonations) is also growing more acute; in one recent test, 48% of video call participants believed an AI agent was a live human. In 2025 alone, the FBI Internet Crime Complaint Center received over 22,000 cybercrime complaints with an AI-related descriptor, with reported losses nearing $900 million. In April of this year, an advanced AI model reportedly uncovered thousands of previously undetected software security flaws, including in a widely used and well-defended open source operating system, and in July, a group of autonomous AI agents worked together to hack a company's systems.

We have previously highlighted the critical importance of public company boards maintaining focus on cybersecurity risk oversight, including developing an understanding of the particular risks each company faces given the systems it uses and the data it holds. Beyond the clear benefit of robust risk oversight to the corporation and its constituents, under the *Caremark* line of Delaware cases, directors may face liability for breach of fiduciary duty only in the exceptional circumstances in which they utterly fail to implement a board-level reporting or information system or consciously fail to monitor such a system or respond to red flags. Such a claim requires a showing of bad faith; ineffective or unsuccessful oversight alone is not enough. But courts have allowed stockholder claims against directors to proceed to discovery where the complaint alleges such failures with specificity.

The accelerating adoption and power of AI tools heightens the imperative for effective board oversight, and boards should consider whether existing reporting-and-escalation structures adequately capture AI-related cyber risks. While the appropriate measures will depend on a company's size, industry, regulatory environment, data assets, AI uses, and threat profile, best practices for cybersecurity oversight in the AI age include the following:

- ***Robust and current data governance***. Boards should oversee a framework that identifies the company's most critical data, maps where it resides, tracks how AI systems access or use it, and assigns clear accountability for its protection across the organization.
- ***Monitor, test, and adapt to the rapidly evolving threat landscape***. Management should be equipped to track developments in AI-enabled cyber threats and defenses, drawing where appropriate on recognized frameworks such as the NIST Cybersecurity Framework and on qualified outside experts. Components of a robust control program may include predeployment and adversarial testing, continuous monitoring, and refinement of vulnerability remediation, containment, and recovery capabilities. Boards should receive periodic reports on material developments, such as significant incidents, testing results, remediation status, and the effectiveness of mitigation efforts.
- ***Oversight of critical AI vendors***. Management should assess the cybersecurity, resiliency, and data-handling practices of key AI service providers, and boards should understand

If your address changes or if you do not wish to continue receiving these memos, please send an e-mail to Publications@Wachtell.com or call 212-403-1443.

where the company is significantly dependent on, or exposed to, those providers and what protections and contingency plans are in place.

- ***Govern risks posed by corporate AI agents***. Boards should understand in general terms what the company's AI agents are authorized to do, what controls limit their access to systems and data, how those controls are tested, and when human approval is required before they take significant actions.
- ***Implement tailored training***. Personnel at all levels of seniority should understand permitted AI uses and applicable safeguards, and undergo training (which may include periodic simulations) to identify and flag deepfakes and other potential threats.

Martin Lipton
Kevin S. Schwartz
David M. Adlerstein