5 KEY TAKEAWAYS

AI and Cybersecurity: Evolving Threats and Risk Mitigation

At the 2026 Kilpatrick Townsend Intellectual Property Seminar (KTIPS), **Greg Silberman** presented **"AI and Cybersecurity: Evolving Threats and Risk Mitigation,"** focusing on how AI is changing the cyber threat landscape and what organizations can do to address emerging risks. The discussion covered AI-enhanced social engineering, accelerated cyberattacks, frontier-model risk, prompt injection, shadow AI, agentic systems, and practical controls for AI governance, architecture, and incident response.

Greg provides these key takeaways:

1

AI Changes the Speed and Scale of Cyber Risk, Not the Fundamentals. AI enables adversaries to operate faster, at greater scale, and with greater personalization and autonomy. Most AI-enabled attacks, however, still exploit familiar weaknesses such as compromised identities, excessive permissions, software vulnerabilities, and human trust. Organizations should continue to strengthen foundational cybersecurity controls while adapting detection, escalation, and decision-making processes to a threat environment that increasingly operates at machine speed.

2

Know Where AI Is and What It Can Do. Maintain an enterprise inventory of AI systems, including AI capabilities embedded in existing software and unauthorized or "shadow AI." Classify risk based on both the sensitivity of the data a system can access and the authority it has to take action. Systems connected to sensitive information, credentials, source code, financial processes, production environments, or internal repositories warrant heightened scrutiny. The key questions are what the AI can see, what it can do, who can control it, what it logs, and how it can be contained.

3

Treat AI Agents as Privileged Nonhuman Identities. Agentic AI can move beyond generating content to access data, call tools, use credentials, modify records, and execute transactions. As an AI system gains authority, permissions, and autonomy, the consequences of misuse or compromise increase. Effective governance and technical controls are therefore essential to the safe and successful deployment of AI agents. Organizations should apply least-privilege access, segregated service accounts, short-lived credentials, tool and API restrictions, and human approval for consequential or irreversible actions. They should also maintain comprehensive logging and the ability to suspend, isolate, or roll back agent activity. Without appropriate controls, a powerful AI agent can quickly create risks comparable to those posed by an insider threat.

4

Treat External Content as Untrusted Input. Prompt injection can turn ordinary emails, PDFs, webpages, support tickets, and other content into malicious instructions that manipulate an AI system. The risk is particularly significant when AI combines untrusted content with enterprise data or action-capable tools. Separate trusted instructions from untrusted content, constrain what retrieved material can influence, apply role-based and purpose-limited access to connected data, validate outputs, and require human approval before high-impact or state-changing actions.

5

Extend Cybersecurity Governance and Incident Response to AI. Organizations do not need an entirely separate cybersecurity program for AI, but existing governance, vendor management, identity controls, secure development, and incident-response processes must account for AI-specific attack surfaces and evidence. Incident plans should address prompts, retrieved content, model and version information, tool calls, approvals, and agent activity, and should establish procedures to suspend or isolate affected AI functionality. Tabletop exercises should include scenarios involving prompt injection, AI-assisted fraud, sensitive-data disclosure, and unauthorized agent actions.

Kilpatrick's AI, privacy, and cybersecurity attorneys can help organizations assess emerging AI-related cyber risks, develop practical governance and technical control frameworks, strengthen vendor and contracting practices, and update incident response plans to address AI-enabled and AI-specific threats.

For more information, please contact:
Greg Silberman,
gsilberman@ktslaw.com