

The Convergence of AI and Security: A Defining Challenge for the Next Five Years

AI's growing autonomy brings new security risks, including cyberattacks, supply chain vulnerabilities, and regulatory challenges. Organizations must adopt zero-trust models, enhance transparency, and strengthen defenses.

BY BRIAN BETKOWSKI AND ED HAINES

As artificial intelligence (AI) reshapes industries and upends traditional business models, it is unlocking opportunities and exposing vulnerabilities. And with the risks tied to poorly secured AI systems mounting, companies and governments alike are being forced to focus on the intersection of AI and security—a convergence that will inevitably shape the future of innovation and strengthen (or weaken) public trust in technology.

But what are the risks? How is the AI landscape evolving? And what can organizations do to navigate these challenges securely? This article will attempt to answer these questions and examine how businesses, policymakers, and security experts can collaborate to mitigate threats, establish safeguards, and ensure that AI-driven advancements remain both innovative and secure.

The Shifting Landscape: Five Defining Trends

The first wave of AI innovation, which began in the mid-20th century with early symbolic AI and rule-based systems, revealed its potential; now, as AI becomes more powerful and ubiquitous, the focus shifts to mitigating its risks. The second wave, which took off in the early 2010s with deep learning and neural networks, ushered in AI's widespread adoption. Now, we are entering a third wave, marked by AI's increasing autonomy and real-world integration. **Five key trends** are shaping the future of AI security, each with profound implications for data protection, trust, and technological stability.

Trend 1 – The Rise of AI-Enabled Cyberattacks

AI's power to analyze and adapt is a double-edged sword. While it drives innovation, it also empowers malicious actors to execute highly sophisticated cyberattacks. Darktrace, a UK-based cybersecurity firm specializing in AI-driven security solutions, has highlighted real-world examples of generative AI conducting sophisticated phishing campaigns. In fact, it has detected AI-powered phishing emails that mimic human writing styles with incredible accuracy, tricking even the most vigilant readers. The challenge for businesses will be building defenses—such as real-time AI-driven threat detection, behavioral analytics to identify anomalies, and multilayered authentication protocols—that are as dynamic as the threats themselves.

Trend 2 – A Global AI Arms Race

The competition to dominate AI technology has escalated into an arms race between nations, with countries increasingly deploying AI to bolster their cyber capabilities and create vulnerabilities in their enemies' critical infrastructure, financial systems, and even national defense. For example, the U.S. Department of Defense's Project Maven, which launched in 2017, uses AI to analyze drone footage to identify enemy combatants, detect movement patterns, and enhance situational awareness to give the military a technological edge. However, such initiatives also raise concerns about how adversaries might develop countermeasures or launch AI-driven cyberattacks on these systems. If governments are willing to take shortcuts in a race to outpace their rivals with AI technology, it could lead to increased security risks and expose gaps that adversaries are quick to exploit.

Trend 3 – Supply Chain Security Under the Microscope

The reliance on third-party AI providers introduces hidden risks into supply chains, making security a shared responsibility across industries and ecosystems. A single breach at a prominent vendor such as Microsoft Azure AI or OpenAI could compromise sensitive systems across a sizeable number of organizations. From open-source AI models to enterprise-grade platforms like Amazon SageMaker and Google Vertex AI, every layer of the supply chain is a potential target. Over the next five years, end-to-end supply chain security will become a priority, requiring companies to rigorously vet their partners and audit every link in the chain.

Trend 4 – The Growing Weight of Regulatory and Ethical Pressures

As AI systems collect and process an unprecedented amount of data, governments are scrambling to impose regulatory frameworks to protect privacy and prevent misuse. As a result, global businesses face a fragmented regulatory landscape with disparate rules across varying regions that complicate compliance. The General Data Protection Regulation (GDPR) in the European Union, for example, has forced U.S. companies like Google and Facebook to retool their AI advertising models to ensure compliance. Noncompliance has led to hefty fines, most notably Google's $57 million penalty in 2019. Navigating these new regulations will require businesses to be agile and demonstrate commitment to the ethical use of AI, as those that fail to align with regulations risk financial penalties, reputational

damage, and operational setbacks.

Trend 5 – The Evolution of Trust in AI Systems

Trust is emerging as the most valuable currency in the AI age. Stakeholders, whether customers, partners, or regulators, will increasingly evaluate AI systems based on their transparency. Companies that fail to demonstrate how their systems protect data, make decisions, and mitigate risks will lose their competitive edge. IBM is a good example of a company that has made significant strides in "explainable AI" with its Watson platform, which offers tools like AI FactSheets to enhance transparency. These FactSheets document AI model performance, fairness, and compliance, helping enterprises build trust with users and regulators. Initiatives like this aim to increase stakeholder trust by ensuring ethical practices and detailing how decisions are made.

Conversely, opaque or unreliable AI models introduce significant risks. For example, Deep Seek, an AI system that has been criticized for generating misinformation and unverifiable claims, exemplifies how a lack of transparency can erode trust. If organizations deploy models that hallucinate data, lack proper safeguards, or fail to disclose decision-making processes, they risk damaging their credibility and exposing users to harmful consequences. In contrast, OpenAI has placed significant emphasis on improving its models' reasoning capabilities to enhance trust and usability. By refining its ability to provide logical, context-aware responses and reducing hallucinations, OpenAI demonstrates how strong reasoning can reinforce transparency. Users and regulators alike are more likely to trust AI when it can articulate the rationale behind its conclusions. In the coming years, trust will separate leaders from laggards in the AI-driven economy.

Strategic Imperatives: Securing the Future of AI

To thrive in this new environment, it's clear that organizations must make AI security a foundational element of their cybersecurity, compliance, and risk management strategies. Failing to secure AI systems doesn't just lead to isolated incidents, it creates systemic vulnerabilities that ripple through industries, economies, and societies. Poor security can result in massive financial losses, reputational damage, and the theft of intellectual property.

For example, last year, MOVEit, a popular file transfer software used by businesses and governments, was compromised in a massive breach. Hackers exploited a vulnerability to steal data from hundreds of organizations, including financial institutions and healthcare providers, exposing millions of records. Simply put, private-sector AI-driven data breaches, which already average millions of dollars in costs, can cripple consumer trust and erode market confidence.

The consequences go beyond financial costs—they threaten human safety and global stability as governments, too, face enormous stakes. Poorly secured AI in critical infrastructure or military systems could lead to sabotage, espionage, or even geopolitical instability. A compromised AI system in government could destabilize democracy or help erode citizens' trust in their leaders. As AI becomes embedded in more industries, from healthcare to transportation to defense, the stakes of poor security grow exponentially.

What's at Stake for Companies? A Breach of Trust and Business Viability.

A single breach can expose sensitive data, cripple critical systems, and erode consumer trust—threatening not just finances but the very viability of a company. From data theft to regulatory

threatening not just finances but the very viability of a company. From data theft to regulatory penalties, the consequences of weak AI security are severe and far-reaching. They include:

Massive Financial Losses

1. **Data Breaches:** Poor security around AI systems can expose sensitive customer data or proprietary algorithms, leading to fines, lawsuits, and loss of business. For example, the average cost of a data breach reached $4.45 million in 2023, and AI-driven attacks are likely to make this figure climb higher.
2. **Operational Disruptions:** Attacks on AI-powered critical systems, such as supply chain management or customer support, could bring operations to a grinding halt, leading to revenue losses and delayed recovery.

Reputation Damage

1. **Trust** is a non-negotiable currency in the digital age. A single high-profile breach could tarnish a company's brand for years. Consumer-facing companies, in particular, risk losing customers who demand transparency and security in how their data is used.
2. For AI providers, failing to safeguard security erodes **credibility**, potentially losing large contracts with enterprises or governments.

Intellectual Property Theft

1. AI systems often house proprietary algorithms, training data, or business-critical insights. A breach could expose this **intellectual property** to competitors or bad actors, undercutting a company's market advantage.
2. State-sponsored hackers targeting R&D-heavy industries like pharma, aerospace, and technology could siphon off years of **innovation** in mere hours.

Regulatory and Legal Consequences

1. Governments worldwide are cracking down on data misuse and breaches, levying **massive fines** on companies that fail to protect sensitive information. For instance, under the General Data Protection Regulation (GDPR), penalties can reach up to €20 million or 4% of annual turnover, whichever is higher.
2. Poor security could also lead to **litigation** from consumers, partners, or shareholders, further compounding financial and reputational fallout.

Compromised AI Models

1. Data poisoning or adversarial attacks can **skew AI systems**, leading to flawed outputs. For example, an AI that mismanages inventory in retail or gives erroneous financial advice could directly harm business operations and lead to loss of trust.

What's at Stake for Governments? National Security and Economic Risks.

From daily government operations to financial transactions, infrastructure management, and military defense, security has never been more critical. Poorly secured systems present significant vulnerabilities, opening the door for adversaries to exploit weaknesses with potentially devastating consequences. Here are some of the key risks to consider:

Infrastructure Sabotage

1. Many governments are deploying AI for critical infrastructure operations, including power grids, water supply systems, and transportation. Poor security in these systems could enable bad actors to **execute catastrophic attacks**, causing widespread disruptions and endangering lives. For example, an attack on a smart city's AI-driven systems could paralyze an entire urban area, impacting millions of residents.

Weaponization of AI

1. Weak security could allow adversaries to **gain control** of military AI systems, including autonomous drones or surveillance platforms, potentially turning these tools against the deploying nation.
2. Even worse, compromised systems could lead to **accidental escalations** or **geopolitical instability** if they act outside human control.

Cyber Espionage

1. Poor security in AI systems used for intelligence gathering could **expose sensitive national secrets**. State-sponsored actors are already exploiting vulnerabilities in AI models to conduct cyber espionage, threatening geopolitical stability and national defense strategies.

Erosion of Public Trust

1. Governments that fail to secure their AI systems risk **losing the trust of their citizens**. For example, a breach of AI systems managing public benefits or healthcare data could result in public outrage, eroding confidence in the government's ability to safeguard personal information.
2. **Loss of trust in elections** or digital governance systems could destabilize democracies, further exacerbating societal divides.

Economic Fallout

1. Inadequate security could disrupt trade, cripple critical industries, or lead to large-scale intellectual property theft, **impacting national economies.** For instance, if a nation's AI-powered financial systems were compromised, the resulting instability could trigger a global economic crisis.

So, what can businesses and governments do to stay ahead of these growing risks? **Here are six key strategies:**

Strategy 1 – Adopt Zero-Trust Security Models

As AI systems collect and process an unprecedented amount of data, governments are scrambling to impose regulatory frameworks to protect privacy and prevent misuse. As a result, global businesses face a fragmented regulatory landscape with disparate rules across varying regions that complicate compliance. The General Data Protection Regulation (GDPR) in the European Union, for example, has forced U.S. companies like Google and Facebook to retool their AI advertising models to ensure compliance. Noncompliance has led to hefty fines, most notably Google's $57 million penalty in 2019. Navigating these new regulations will require businesses to be agile and demonstrate commitment to the ethical use of AI, as those that fail to align with regulations risk financial penalties, reputational damage, and operational setbacks.

Strategy 2 — Develop AI-Specific Threat Models

Traditional cybersecurity measures, such as firewalls and endpoint protection, are insufficient to protect AI systems. Companies need to account for specific risks like data poisoning and tailor their security approaches to the specific vulnerabilities of AI. For instance, Microsoft has developed an AI threat matrix to map out attack vectors specific to machine learning models, helping organizations preemptively identify security gaps.

Strategy 3 — Invest in Explainability and Transparency

As trust becomes paramount, explainable AI (XAI) will play a critical role in reassuring stakeholders. Companies that make their models' decision-making processes clear and accessible will build stronger relationships with customers, clients, and regulators. One example is Hugging Face, a New York–based AI model repository that provides transparency through "Model Cards." These documents disclose key details about an AI model's strengths, weaknesses, training data, and intended uses. By providing this level of insight, Hugging Face helps developers and businesses understand potential biases, risks, and limitations before integrating AI models into production environments. Similarly, organizations can adopt "AI FactSheets," a framework developed by IBM, which outlines AI model performance, fairness, and compliance to enhance transparency. In regulated industries like healthcare and finance, where AI-driven decisions impact lives, such transparency measures are essential for maintaining public trust. By investing in explainability and transparency, businesses can mitigate regulatory scrutiny, reduce bias-related risks, and foster greater adoption of AI-driven solutions.

Strategy 4 — Rigorously Vet AI Vendors

A secure AI ecosystem begins with thorough due diligence. Organizations must ensure that their AI providers meet rigorous security standards, conduct regular audits, and comply with evolving regulations. One easy way to do this is by requiring vendors to adhere to established security frameworks such as NIST's AI Risk Management Framework or ISO/IEC 42001, which provide clear guidelines for AI security.

Strategy 5 — Build Collective Defense Mechanisms

Collaboration will be key to addressing AI security challenges. Businesses and governments should participate in information-sharing coalitions, align with best practices, and coordinate responses to emerging threats. For example, the Cyber Threat Alliance (CTA) is a Virginia-based nonprofit coalition of cybersecurity companies, including Palo Alto Networks and Fortinet, which shares real-time threat intelligence. This initiative has enabled member organizations to detect and mitigate AI-driven threats faster than individual companies could on their own. Another strong example is NATO's Cooperative Cyber Defense Centre of Excellence (CCDCOE), which fosters international collaboration on cyber defense. By pooling expertise and resources across member nations, CCDCOE strengthens global cyber resilience against AI-powered threats, such as deepfake propaganda and automated cyberattacks. As AI-driven threats grow, collective defense mechanisms will become essential. The more organizations share insights and best practices, the better prepared they will be to detect, prevent, and mitigate AI-related security risks.

Strategy 6 — Foster a Security-Conscious Culture

Every employee in an organization, from IT and HR to accounting and the C-suite, must understand

Every employee in an organization, from IT and HR to accounting and the C-suite, must understand the risks AI systems pose. Regular training and simulations can cultivate a culture of vigilance, empowering organizations to identify and respond to potential breaches more effectively. Netflix provides a great example of fostering a security-conscious culture through its use of "Chaos Monkey," a tool that deliberately disrupts systems to test resilience. By simulating failures and vulnerabilities, Netflix ensures that its employees are continuously prepared to respond to security threats. Similarly, companies can conduct "red teaming" exercises where cybersecurity experts simulate attacks on AI systems to expose weaknesses before malicious actors exploit them. Regular phishing awareness training, penetration testing, and real-world cyber incident simulations help employees at all levels recognize potential threats and take proactive security measures. AI security isn't just the responsibility of IT teams; it must be ingrained into the entire organizational culture to protect against evolving threats. Companies that prioritize security awareness and preparedness will be far better equipped to prevent breaches and minimize risks.

What's Next?

In the next five years, the convergence of AI and security will define the leaders of the digital age. Companies and governments must recognize that AI security isn't just a cost center, it's a strategic differentiator. Those who build resilient, trustworthy systems will earn the confidence of stakeholders and seize opportunities in an increasingly AI-driven world.

The challenge is daunting, but the rewards of bold, secure innovation are immeasurable. By prioritizing security as a foundation for AI deployment, organizations can turn risks into opportunities and shape a future where innovation and safety coexist.

VIEW → SPRING 2025 TABLE OF CONTENTS

Subscribe to the Journal

Sign Up

The Jabian Journal is a publication of **Jabian Consulting** | **Careers** | **About Us**