KPMG

AI value depends on AI security

A guide for CISOs to assess and manage risks in AI ecosystems

kpmg.com/us

Getting up to speed on AI security

Artificial intelligence (AI) is redefining the business world. This game-changing technology is fast becoming a competitive differentiator, and even table stakes in some cases. In a 2023 KPMG survey of technology professionals, 57 percent said that AI and machine learning will be important in helping them achieve their business objectives over the next three years. In addition, 68 percent said these technologies will be vital in helping them to achieve their short-term business goals.[1]

By designing, building, and deploying AI solutions at scale, organizations can help support better decision-making, identify new opportunities for growth, help reduce operational costs, and improve customer experiences. But as businesses integrate AI at scale, existing risks are amplified and new risks are clearly present. In another recent KPMG survey, 63 percent of business leaders considered cybersecurity and personal data privacy as priorities for risk management.[2]

As the presence of AI expands and AI capabilities rapidly evolve, chief information security officers (CISOs) and their security teams face a multifaceted challenge. A key issue is quickly assessing the range of risks AI technology poses across the enterprise. Another is embedding the right framework to help all business stakeholders effectively manage AI risks and secure the enterprise. CISOs also need to help ensure their company's AI solutions are compliant with rapidly evolving standards and regulations, both in the US and overseas.

---

[1] "KPMG global tech report 2023," KPMG International, 2023

[2] Findings are based on KPMG surveys on generative AI in 2023. Data points are based on 225 U.S. respondents.

What you'll learn

Organizations are looking to move from strategy and planning to the operationalization of secure AI ecosystems. In line with this objective, the report provides information, insights, and suggestions that can help CISOs better assess and manage security risks due to enterprise-wide AI technology. We discuss:

Critical AI-related security risks

Elements of a leading-edge AI security framework

Benefits of enhanced AI risk management

Leading practices for CISOs to enhance AI security

How KPMG can help CISOs develop AI security solutions

Understanding security risks caused or exacerbated by increased AI adoption

AI is changing the threat landscape for today's organizations.

Unlike traditional software where algorithms follow a set learning path, AI applications involve an iterative process based on autonomous learning and a stream of data from both within and outside of the organization to support large language models (LLMs). This process is a key part of the power of AI: It enables AI systems to learn and grow in real time, support the automation of repetitive tasks, and process large amounts of data to find key patterns and anomalies.

However, some characteristics of AI applications—their ability to learn and change, automate complex processes, and leverage huge amounts of data—can also introduce critical risks that could expose the organization to major harm, such as data breaches, financial loss, and reputational damage. Examples include the following:

Data poisoning and bias: These are corrupting model-training data sets with deliberately flawed inputs that compromise data integrity. An attacker could add malicious data to a training data set to bias the model toward a specific classification or conclusion.

Breaches of sensitive data: Breaches extract data directly out of a model trained on data from the internet that contains sensitive information, including personal information and intellectual property (IP).

Prompt injection: This tricks an AI tool into bypassing its normal restrictions. It is done by using prompts that can override the controls that define how and by what rules AI applications interact with users. Prompt injections can also fool the system into thinking it does not need to follow those rules anymore.

Model theft and replication: The output of a model is used to infer some of its parameters or architecture, also known as model inversion. These inferences can then be used to steal the model by creating a copy or to extract sensitive information that was used to train the model.

Model evasion: This forces the model to classify incorrectly during the inference (or prediction/post-training) phases. Sometimes similar to poisoning attacks, model evasion is used by adversaries looking to avoid detection systems such as spam filtering and malware detection.

Backdoor attacks: A hidden entry point is created into hardware, software, or networks that can be exploited by an attacker to gain unauthorized access. Backdoors can be created intentionally by attackers or unintentionally by developers. They can be used to steal sensitive data, install malware, or carry out other malicious activities.

Phishing: Attackers generate highly personalized phishing emails by crawling social media platforms and the internet with AI tools, looking for user-generated content and other public information about individuals.

Trojan attacks: These attacks insert malicious code into a model during the training process. The goal of a Trojan attack is to create a backdoor in the model that can be exploited by an attacker to perform malicious actions.

Regulatory penalties: Penalties result when organizations are not keeping pace with rapidly emerging AI regulations and standards. The EU's AI Act, for example, will require providers of many AI systems to implement several governance mechanisms. Various AI controls and standards also need to be addressed, including the MITRE ATLAS framework, NIST AI RMF 1.0, Google's Secure AI Framework, and the Microsoft Responsible AI Standard.

AI tools to fight AI risks

AI technology can be used by organizations to support their cybersecurity defenses based on capabilities such as the following:

Real-time threat detection: Continuously analyze data and network activity to identify abnormalities and deviations from normal behavior, helping to quickly detect potential cyber intrusions and thereby mitigate the impact of cyberattacks.

Continuous monitoring and alert enrichment: Free up the cybersecurity team for other tasks and reduce the burden of excessive false alarms and alert fatigue. AI automation also gives analysts additional information to support more accurate threat assessments and improved decision-making regarding attacks.

Data leakage protection: Rapidly classify high-risk data and uncover patterns in behavior around data to detect anomalies and reduce delays in remediation. Also control and block data transfers that violate policies related to data access and data handling.

Automated documentation and improved communication: Generate accurate and concise documentation as well as reports and explanations translated from complex technical terms to easy-to-understand content. This accelerates the documentation process and helps to ensure that important information is readily available to key stakeholders.

Phishing detection: Use pattern recognition and analysis abilities to quickly and accurately identify phishing emails and differentiate them from spam and legitimate emails.

Compliance support: Use AI to forecast potential compliance risks based on historical data. AI can also support document analysis by quickly reviewing large amounts of legal and regulatory documents to help demonstrate compliance.

Developing a thorough AI security framework

AI requires new ways of thinking about security and privacy. In traditional IT environments, security management involved switching out hardware components and releasing new software versions or patches on a regular basis to address new threats. Today's AI systems are far more dynamic, interactive, and customized, with more risks and the need to address these risks in real time.

Accordingly, AI security risk management must be proactive, not an afterthought. Security must be built into AI systems from the ground up, embrace the entire AI environment, and be continuously monitored for safety, security, and proper governance.

A thorough, holistic AI security framework can help address these goals based on the following activities:

- The development of an enterprise-wide framework for assessing and securing AI systems, backed by policy and procedure documents in line with security and regulatory requirements
- A survey of existing AI platforms, data sources, tools, and architectures
- The identification of where AI is currently being used in the organization and by whom
- A review of existing AI policies and procedures
- A mission statement in alignment with organizational values that outlines the goals and commitments of AI security initiatives
- The identification of resource and capacity requirements, goals and metrics, reporting structures, and response plans
- Evaluation of future regulations and security issues.

Benefits of enhanced AI risk management

By its nature, AI is a complex, fluid, and rapidly evolving technology. New security threats and issues are emerging on a regular basis. But with a well-designed AI security framework in place, organizations can better seize the value of AI in a quick, confident, and responsible manner.

Here are some potential security and business benefits of implementing a thorough AI security framework:

1. **Improved collaboration between data science and cyber teams.** In many organizations, data science and cyber teams don't work closely together. Collaboration can help define who bears responsibility for security, ethics, and governance; increase rigor around budget and resourcing; and support a consistent approach in protecting IP.

2. **Enhanced cyberattack monitoring, detection, and response.** AI technology increases the number and variety of attack vectors for cybercriminals. Improved risk management not only can help secure AI systems but also can enhance the overall security posture of the organization through the detection of and response to cyberattacks.

3. **Increased visibility and control over the AI data lifecycle.** AI models that ingest IP, trade secrets, personal identifiable information, or protected health information aren't always created with safeguards that protect against data loss or corruption. AI's ability to detect patterns and synthesize structured and unstructured data can support improved visibility and control over the data lifecycle.

4. **Expanded transparency into third-party vendors' AI use.** Bad actors often use third-party vendors and other entities in supply chains as a gateway into an organization's IT environment for ransomware attacks, the theft of IP and personal credentials, and other attacks. AI can support greater transparency into vendor interactions and their black box software.

5. **Greater integration of regulatory requirements.** The current regulatory environment is changing on a regular basis. AI tools—when properly secured and governed—can help organizations review regulatory issues, provide documentation, and help demonstrate compliance.

6. **Acceleration of AI value. In some ways, AI applications are relatively easy to set up and use.** This means that new AI algorithms are emerging every day, often without sufficient security safeguards. With proper risk management, organizations can safely move forward with grassroots and enterprise AI initiatives, enabling them to leverage the strengths and benefits of AI without increasing exposure to security risks.

Three leading practices for CISOs to help ensure enterprise AI security

Drive cross-functional collaboration and coordination: Implementing secure AI applications is an enterprise-wide issue, demanding skills, perspectives, and responsibilities from professionals throughout the business. Develop a framework for collaboration on innovation and secure AI. Stand up a cross-functional steering committee comprising data science teams, security, legal, privacy, digital, etc. And always loop in the risk management function as new AI-based products are greenlit.

Hire a chief trust officer: As AI is integrated into more products and processes, the chief trust officer is a new role for businesses to consider. The chief trust officer is responsible for helping ensure there is trust not only in data, but also in humans working on AI projects. Companies such as SAP, Cisco, IBM, OneTrust, Microsoft, and Salesforce have led the way in establishing this executive-level position, many by promoting former CISOs.[3]

Develop AI-focused policies and controls: Identify where AI is being used, including small projects, independent contractors, and suppliers. Develop a responsible-use policy that governs your use of AI. Establish controls to adhere to that policy. And develop an automated and efficient approach to monitor and remediate the controls.

[3] "The Chief Trust Officer Role Can Be the Next Career Step for CISOs," InformationWeek, November 14, 2022

How KPMG can help

KPMG Cyber Security Services is a leading AI security and trust service provider, developing and delivering effective security solutions for AI systems and models across multiple industries. AI security services are designed to help organizations assess their AI ecosystem, secure their critical models, and respond to adversarial attacks. Our risk-based approach provides effective prioritization to secure an organization's most critical systems.

The KPMG AI security framework design is central to our offerings, providing security teams with a tested playbook to proactively assess their organization's AI systems in development and production environments. The framework helps to secure those systems against threats and support a fast, effective response to attacks. Services include the implementation of a broad suite of AI security tools. The framework also supports AI red teaming where professionals conduct prearranged, structured testing to find flaws and vulnerabilities in systems. We tailor the approach to meet the requirements, platforms, and capabilities of different organizations, helping to deliver an effective and accepted security strategy.

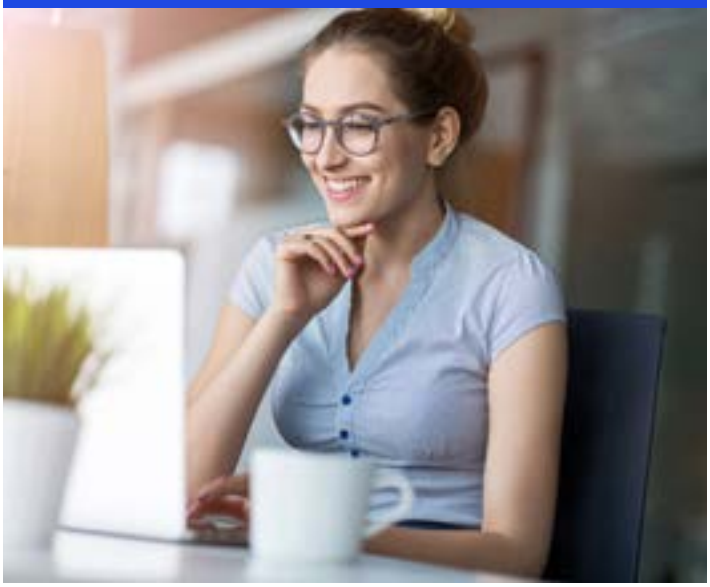
We also offer deep experience in regulations, risk, security, privacy, and other critical areas that can prove beneficial in the fast-emerging space of trusted AI. Our powerful network of strategic alliances and investments enhance our integrated capabilities and help our clients seize more value from strategy and technology investments. In addition, KPMG has launched a cyber center of excellence (COE) for securing AI to help support the development of our AI security framework, as well as other services to help organizations prepare for and better manage adversarial AI cyber threats.

For more information, visit:
https://kpmg.com/us/en/capabilities-services/advisory-services/cyber-security-services.html.

Cranium: Building trust in AI with KPMG

Cranium is a cutting-edge software company supported by KPMG that enables organizations to secure their AI technologies. The Cranium Enterprise software platform was developed in collaboration with KPMG AI security specialists in KPMG Studio, our firm's start-up incubator. Serving as a technology enabler for the KPMG AI security framework, the Cranium platform helps cybersecurity and data science teams to understand where and how AI is impacting their systems, data, or services, all without interrupting workflow.

For more information, visit:
https://www.cranium.ai/.



Get in touch

Matthew P. Miller
Principal,
Cyber Security Services
KPMG LLP
E: matthewpmiller@kpmg.com

Katie Boswell
Managing Director,
Cyber Security Services
KPMG LLP
E: katieboswell@kpmg.com

Some or all of the services described herein may not be permissible for KPMG audit clients and their affiliates or related entities.

Learn about us: **kpmg.com**