



Home



Introduction



Related Summit  
Sessions



Contributors



User  
Guide



Securing AI



Risk &  
Compliance



Policy &  
Governance



AI Bill of  
Materials



Trust  
& Ethics

# Balancing ROI and Risk

A **Practitioners' Guide** to Managing AI Security





Home



Introduction



Related Summit  
Sessions



Contributors



User  
Guide



Securing AI



Risk &  
Compliance



Policy &  
Governance



AI Bill of  
Materials



Trust  
& Ethics

# Introduction

## Dear Readers,

Companies of all sizes and across industries are engaged in the Artificial Intelligence (AI) revolution. The race to integrate AI into internal operations and bring AI-based products and services to market is moving faster than almost anyone could have imagined. These technologies stand to help companies transform their businesses, achieve short- and long-term objectives at an historic pace, and drive deeper connections with customers, partners, and other stakeholders.

At the same time, the fervent excitement about AI has the potential to relegate critical security and assurance considerations to afterthoughts. Recognizing this disconnect – between AI innovation and AI Security – Global Resilience Federation (GRF) convened an AI Security & Trust working group and asked KPMG to facilitate in-depth discussions between AI and security practitioners from more than 20 leading companies, think tanks, academic institutions, and industry organizations. KPMG was also asked to document the output of the working group sessions, which, ultimately, led to the creation of this guide.

The ***Practitioners' Guide to Managing AI Security*** aims to provide insights and considerations that strengthen collaboration between data scientists and AI security teams across five tactical areas identified by the working group: Securing AI, Risk & Compliance, Policy & Governance, AI Bill of Materials, and Trust & Ethics.

I thank all our participants for their valuable input to this guide, which I hope will help companies secure AI as they uncover and apply the incredible potential of this technology.

**Thank you,**  
**Mark Orsi, CEO**  
**Global Resilience Federation**





Home



Introduction



Related Summit Sessions



Contributors



User Guide



Securing AI



Risk & Compliance



Policy & Governance



AI Bill of Materials



Trust & Ethics

# Upcoming AI Security Summit

## Global Resilience Federation (GRF) Summit on Security & Third-Party Risk

Rapid advancements in Artificial Intelligence (AI) capabilities have instigated an equally quick paradigm shift in how organizations approach processes across business functions. From automation of simple tasks all the way to highly sophisticated models providing diagnostic recommendations based on medical imaging, AI has proven to be an exceptional tool for gaining and maintaining competitive advantage in tumultuous markets. However, even as it becomes evident that AI outputs will likely be critical to the future success and health of companies across industries, threat actors are taking notice of the new attack surface the technology creates.

## The Global Resilience Federation (GRF) Summit on Security & Third-Party Risk being held October 11-12, 2023, in Austin, Texas, will illuminate how AI Security and AI innovation can be pursued in tandem.

The summit will include a joint keynote on Responsible AI from KPMG and Cranium leadership, as well as panels offering unique insights from both AI and cybersecurity leaders and practitioners on ways organizations are managing AI Security across sectors. This critical and engaging conference is designed for CIOs, CISOs and AI/ML experts, who will find value in engaging on effective ways to manage security and trust as they institute Artificial Intelligence and Machine Learning models in their organizations.





Home



Introduction



Related Summit Sessions



Contributors



User Guide



Securing AI



Risk & Compliance



Policy & Governance



AI Bill of Materials















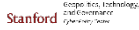



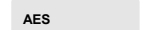
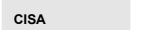



Trust & Ethics

# Contributors

## Our Thoughtful Contributors

KPMG and GRF would like to express our sincerest appreciation for the support of our community members, who have invested their time, thoughts, and energy into an effort to push forward the collective security industry’s approach to the AI revolution. Together, we assembled what we hope you will find to be an insightful and timely guide on how organizations can approach securing AI.

 <b>KPMG</b> Katie Boswell Matthew Miller Kristy Hornland Donna Ceparano John Hodson Kelsey Flynn Riley Richards	 <b>Cranium</b> Jonathan Dambrot Felix Knoll Paul Spicer Daniel Christman	 <b>Campbell Soup Company</b> Martin Bally Mark Wehrle Shelley Ivanko Joshua Vrancik Brian Roberts	 <b>Kenvue</b> Mike Wagner Chris Van Schijndel David Merritt	 <b>Johnson &amp; Johnson</b> Gary Harbison Hal Stern Bill Janicki Michael Barrett	 <b>Cyber Defense Matrix</b> Sounil Yu	 <b>Tampa Electric Company</b> Terri Khalil Mayda Gonzalez	 <b>Stoel Rives</b> Jon Washburn	 <b>Wells Fargo</b> David LaFalce Vamsi Kadiyala Dale Miller Jeff Stapleton Matthew Campbell Peter Bordow Satish Katakam Jerry Flanagan	 <b>Chevron</b> Margery Connor Ashish Shah	 <b>MITRE</b> Christina Liaghati
 <b>Amherst College</b> Scott Alfeld	 <b>Microsoft</b> James Ringold	 <b>Hinshaw Law</b> Sherri Vollick	 <b>Stanford</b> Jim Dempsey	 <b>NYU</b> Joel Caminer Quanyan Zhu	 <b>Shared Assessments</b> Andrew Moyad	 <b>Sharpe Management Consulting</b> Alex Sharpe	 <b>AES</b> Ryan Boulais Sean Otto	 <b>CISA</b> Garfield Jones Christine Lai	 <b>Paul   Weiss</b> James Forrest	





Home



Introduction



Related Summit Sessions



Contributors



User Guide



Securing AI



Risk & Compliance



Policy & Governance



AI Bill of Materials



Trust & Ethics

# User Guide

## Overall Structure of this Output:

Each area of consideration (Securing AI, Risk & Compliance, Policy & Governance, the AI Bill of Materials, and Trust & Ethics) was identified as a critical topic on which to provide guidance and considerations by the Global Resilience Federation (GRF) cross-industry roundtable participants. Within each of these sections, which can be accessed directly via the navigation bar at the top of this guide, you will find the following breakouts:



### Overarching Themes

Each area of consideration features three to four central themes that were identified as consistent challenges or experiences among roundtable participants.



### Recommendations

Aligning with the overarching themes, recommendations and better practices identified by roundtable participants are highlighted to assist practitioners in addressing the challenges uncovered.





Home



Introduction



Related Summit Sessions



Contributors



User Guide



Securing AI



Risk & Compliance



Policy & Governance



AI Bill of Materials



Trust & Ethics

# Securing AI

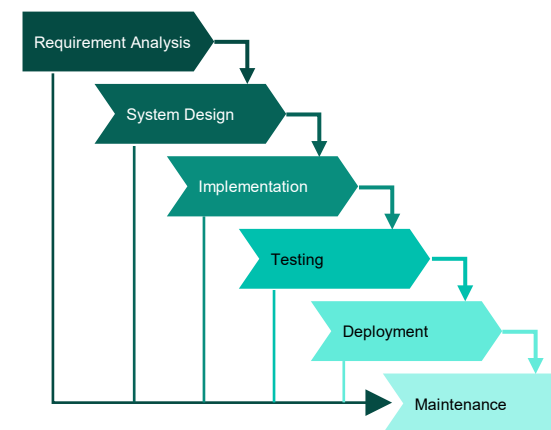
## Scoping the Approach

The multi-industry working group that developed this practitioners' guide quickly determined it was important to define ground rules for what makes Artificial Intelligence (AI) Security different from the traditional approach to cybersecurity.

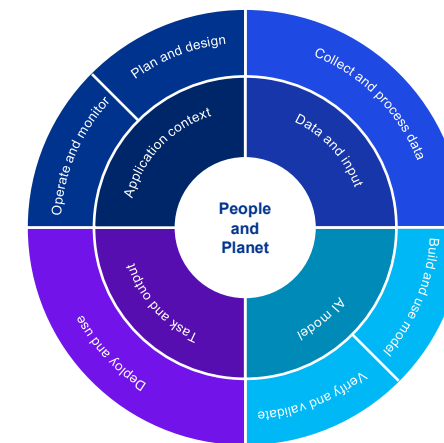
- The first order of business was articulating the difference between traditional software development and AI development lifecycles.
- Traditional software is built using a clear set of deterministic steps, which dictate how the software should operate consistently in production. An apt analogy would be the steps you would take to construct a regulation football.

In contrast, AI models are trained on large datasets to learn patterns and make predictions. These models are much less akin to a static product, and much more like training an elite athlete. As with athletes, AI learns through inputs and can enhance their skills with new patterns for success. And, even once a skill is learned, both athletes and AI require monitoring to ensure they continue desired behaviors.

When it comes to AI Security, it is critical to ensure that both traditional security controls and nuanced controls unique to AI are blended in a holistic approach to safeguarding your organization.



Traditional Software Lifecycle



AI Lifecycle (NIST AI RMF)

“...An AI system is an engineered or machine-based system that can, for a given set of objectives, generate outputs such as predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy...” NIST Artificial Intelligence Risk Management Framework<sup>3</sup>





Home



Introduction



Related Summit Sessions



Contributors



User Guide



Securing AI



Risk & Compliance



Policy & Governance



AI Bill of Materials



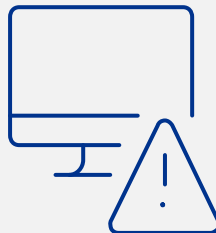
Trust & Ethics

# Securing AI

## Expanded Attack Surface

For practitioners, it is important to determine the types of exploits unique to AI that organizations have begun to model as potential security risks. Click on the categories below to see definitions of four reoccurring threat models.

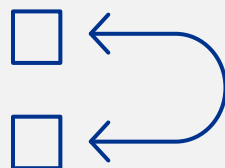
- 
- 
- 



Poisoning



Evasion



Inference



Functional Extraction





Home



Introduction



Related Summit Sessions



Contributors



User Guide



Securing AI



Risk & Compliance



Policy & Governance



AI Bill of Materials



Trust & Ethics

# Securing AI

## Expanded Attack Surface

For practitioners, it is important to determine the types of exploits unique to AI that organizations have begun to model as potential security risks. Click on the categories below to see definitions of four reoccurring threat models.

- 
- 
- 

### Data Poisoning

#### Description

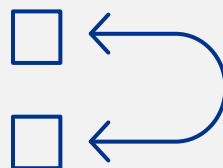
Contamination of ML system training data to get a desired outcome at inference time. With influence over training data, an attacker can create “backdoors” where an arbitrary input results in an undesired output through “reprogramming.”

#### How

An attacker inserts corrupt data into a training dataset to compromise a machine learning model during training. Some techniques aim to trigger a specific behavior in a computer vision system when it faces a specific pattern of pixels at inference time. Others aim to reduce the accuracy of a machine learning model on one or more output classes.



Evasion



Inference



Functional Extraction





Home



Introduction



Related Summit Sessions



Contributors



User Guide



Securing AI



Risk & Compliance



Policy & Governance



AI Bill of Materials



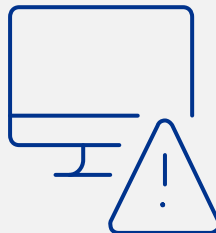
Trust & Ethics

# Securing AI

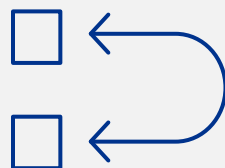
## Expanded Attack Surface

For practitioners, it is important to determine the types of exploits unique to AI that organizations have begun to model as potential security risks. Click on the categories below to see definitions of four reoccurring threat models.

- 
- 
- 



Poisoning



Inference

### Evasion

#### Description

Modification of a query to get a desired outcome. These attacks are performed by iteratively querying a model and observing the output.

#### How

An adversary inserts a small perturbation (noise) into an ML model's input to cause incorrect classification. Although they are similar to poisoning attacks, evasion attacks differ in that they try to exploit weaknesses of the model in the inference phase, not the training phase. The more knowledge an attacker has about your model and how it's built, the easier it is for them to mount an attack.



Functional Extraction





Home



Introduction



Related Summit Sessions



Contributors



User Guide



Securing AI



Risk & Compliance



Policy & Governance



AI Bill of Materials



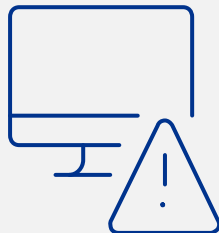
Trust & Ethics

# Securing AI

## Expanded Attack Surface

For practitioners, it is important to determine the types of exploits unique to AI that organizations have begun to model as potential security risks. Click on the categories below to see definitions of four reoccurring threat models.

- 
- 
- 



Poisoning



Evasion

### Inference

#### Description

Recovering the features used to train a model. A successful attack would result in an attacker being able to launch a membership inference attack, which could compromise private data.

#### How

Inference attacks aim to reverse the information flow of an ML model. An adversary gains knowledge of data that was not explicitly intended to be shared. These attacks pose severe privacy and security threats to individuals and systems. Attacks are considered successful if private data are statistically correlated with public data and ML classifiers capture these correlations.



Functional Extraction



Home



Introduction



Related Summit Sessions



Contributors



User Guide



Securing AI



Risk & Compliance



Policy & Governance



AI Bill of Materials



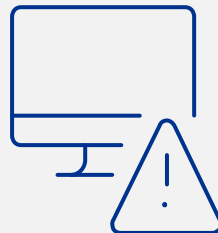
Trust & Ethics

# Securing AI

## Expanded Attack Surface

For practitioners, it is important to determine the types of exploits unique to AI that organizations have begun to model as potential security risks. Click on the categories below to see definitions of four reoccurring threat models.

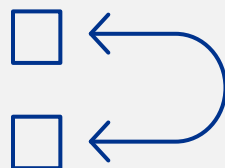
- 
- 
- 



Poisoning



Evasion



Inference

### Functional Extraction

#### Description

Recovering a functionally equivalent model by iteratively querying the model. This allows an attacker to examine the offline copy of the model before further attacking the model online.

#### How

Functional extraction involves making requests to the target model with inputs to extract as much information as possible. With the set of inputs and outputs, the attacker can train a model called a “substitute model.”





Home



Introduction



Related Summit Sessions



Contributors



User Guide



Securing AI



Risk & Compliance



Policy & Governance



AI Bill of Materials



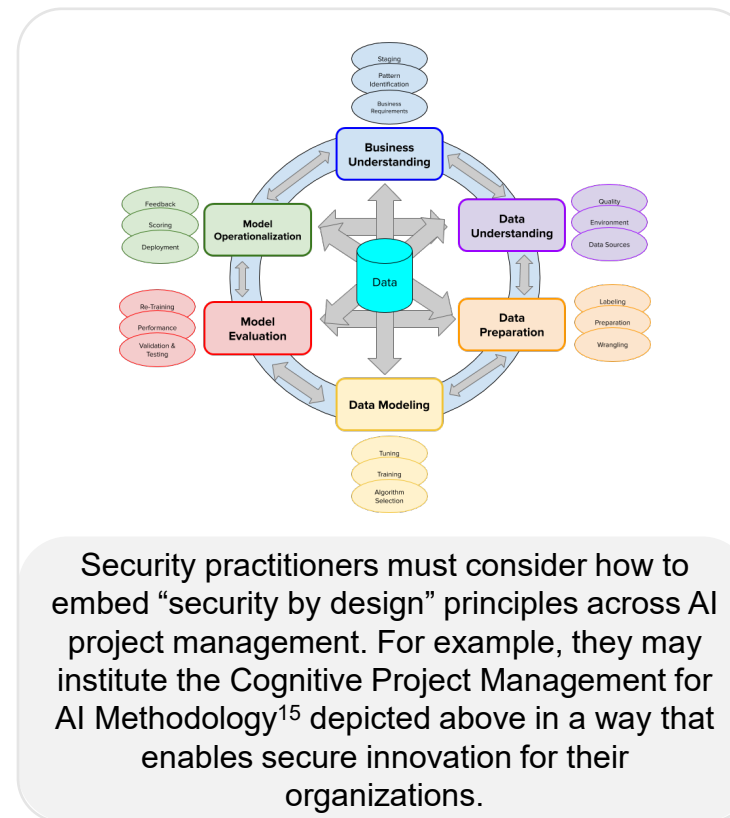
Trust & Ethics

# Securing AI

## Reducing Risk without Jeopardizing Return

While AI use cases across organizations vary from leveraging Large Language Models (LLMs) to virtual agents to machine vision, security practitioners value coordination with both the business and AI / data science teams. This allows organizations to continuously mitigate risk through the unique AI lifecycle without disrupting innovation and operations. In addition, most organizations describe their AI functions as “decentralized,” which can result in teams needing to partner on each individual use case to assess potential threats, risks, and relevant security controls.

The following sections in this guide include Risk & Compliance, Policy & Governance, AI Bill of Materials, and Trust & Ethics, which provide guidelines on the relative challenges and potential responses organizations can consider in their own AI Security journeys. These materials leverage guidance from NIST AI RMF 1.0, the MITRE ATLAS Framework, Microsoft Responsible AI Standard (v2), ISO/IEC DIS 5338, and regulations such as the EU AI Act to support potential strategies and mitigation approaches.





Home



Introduction



Related Summit Sessions



Contributors



User Guide



Securing AI



Risk & Compliance



Policy & Governance



AI Bill of Materials



Trust & Ethics

# Risk & Compliance

## Introduction

As security practitioners are aware, the rise of AI has brought about a paradigm shift in the threat landscape for large enterprise organizations. While AI presents unprecedented opportunities for optimizing operations and driving innovation, it also introduces unique risks and compliance challenges that demand specialized attention. With increasing regulatory scrutiny, managing the risk and compliance of AI systems has become a paramount concern for both security teams and leadership.

Inadequate risk and compliance management for AI systems can result in severe consequences, including financial penalties, reputational damage, legal liability, and loss of stakeholder trust. The intricate and opaque decision-making processes of AI systems pose distinct challenges for traditional risk management and compliance frameworks, necessitating the adoption of specialized better practices.

## NIST AI RMF Core<sup>3</sup> – AI RMF 1.0





Home



Introduction



Related Summit Sessions



Contributors



User Guide



Securing AI



Risk & Compliance



Policy & Governance



AI Bill of Materials



Trust & Ethics

# Risk & Compliance

## Theme #1: Creating a roadmap toward an AI risk assessment framework

Many organizations begin with AI risk management, i.e., the “risk identification” stage. This stage involves defining the organization’s AI threat profile and determining where AI risk fits into overall IT risk management programs.



As organizations build out AI model use cases, they are advised to consider cybersecurity and privacy considerations in tandem, both as they apply to internal AI models and to those introduced by third parties with which they do business. Regarding the latter, a better practice is to require the third party to provide a Bill of Materials to allow the organization to inventory and track potential risks.

Currently, risk identification and threat profiling are not yet very mature at many organizations, although awareness efforts are underway, as well as guidance for establishing formal documentation processes.



## Recommendations

Organizations that are still in the early stages of building AI-specific risk management programs, should stress the importance of reaching maturity as soon as possible. As you develop programs to proactively identify, assess, and mitigate risks associated with AI technologies, consider the following:

- **Develop an AI risk assessment framework:** Establish a formal and structured process for identifying and assessing risks associated with AI systems. This includes understanding the AI threat profile, evaluating cybersecurity considerations, and conducting risk assessments for AI model use cases. Additionally, the framework should include guidelines, checklists, and templates that can be used to assess the risks associated with different stages of the AI lifecycle.
- **Enhance risk documentation and communication:** Focus on formal documentation processes to capture risks associated with AI systems. Develop clear risk profiles and baseline levels of risk for AI solutions across the organization's portfolio. Communicate these risks to relevant stakeholders, including senior leadership, IT teams, data scientists, customers, and others.





Home



Introduction



Related Summit Sessions



Contributors



User Guide



Securing AI



Risk & Compliance



Policy & Governance



AI Bill of Materials



Trust & Ethics

# Risk & Compliance

## Theme #2: Keeping up with regulatory developments

As regulatory considerations for AI systems evolve, security teams need to demonstrate due diligence when it comes to compliance. In the U.S., this may require a strategy to comply with privacy laws that are still new and evolving, as only a few U.S. states currently have robust laws in place.

- 
- 
- Following are some regulations on which an organization should consider focusing today. For example, regulations that govern personal health information (PHI), like HIPAA, are of critical relevance to many organizations, such as law firms and professional services firms that handle clients' medical data. Further, there is the possibility that, if AI is used as part of the financial reporting process, the activities could be subject to controls like the Sarbanes-Oxley Act (SOX) or scrutiny from regulatory bodies like the Securities and Exchange Commission (SEC).
- 

Finally, EU regulations such as the GDPR data subject individual rights, including, the "right to be forgotten," could have clear implications for both training and input data used in AI systems. Other EU regulations may also come into play, even for organizations that are primarily based in the U.S.



## Recommendations

In efforts to remain in compliance with current and potential regulations, consider leveraging these strategies:

- **Stay informed and updated:** It is essential for organizations to stay informed about potential regulations that may impact their AI aspirations and related operations. This can be achieved through continuous monitoring of regulatory developments, engaging with industry experts and legal advisors, and actively participating in relevant industry forums, conferences, and workshops to gain insights.
- **Conduct a comprehensive regulatory assessment:** Conduct a comprehensive assessment of the regulatory environment as it relates to the growing focus on AI globally to understand the specific regulations that are applicable to your industry, geography, and use cases. This assessment should include a review of existing regulations related to data protection and privacy, as well as industry-specific regulations, as well as any pending regulations that may impact the organization's AI initiatives.





Home



Introduction



Related Summit Sessions



Contributors



User Guide



Securing AI



Risk & Compliance



Policy & Governance



AI Bill of Materials



Trust & Ethics

# Risk & Compliance

## Theme #3: Mitigating risk without process interruption

A key focus for organizations starting their AI journeys is mitigating risks associated with AI systems while ensuring that data science teams are not interrupted in their development processes. Data science teams often work in agile and fast-paced environments, experimenting with different data sources, models, and algorithms to iteratively develop and improve AI systems.

Security and privacy practices lie outside of the core competencies of data science and machine learning. Additionally, the AI / ML threat landscape is constantly evolving, with new risks and attack vectors emerging regularly.



## Recommendations

When determining risk mitigation strategies, consider the following mechanisms to avoid creating friction:

- **Integrate security and privacy into the AI lifecycle:** Embed security and privacy considerations as nonnegotiable parts of the AI development lifecycle, starting with the planning and design stages. For example, since scoping AI use cases is a critical part of the development process, embedding security considerations will help account for pertinent risks early in the process. If possible, involve members of the security team in initial design meetings to support the data science team in identifying potential risks.
- **Provide continuous security education and training:** Offer data science teams regular training and education on security and privacy better practices, threat awareness, and compliance requirements. This will enable them to incorporate the relative security measures into their development activities.







Home



Introduction



Related Summit Sessions



Contributors



User Guide



Securing AI



Risk & Compliance



Policy & Governance



AI Bill of Materials



Trust & Ethics

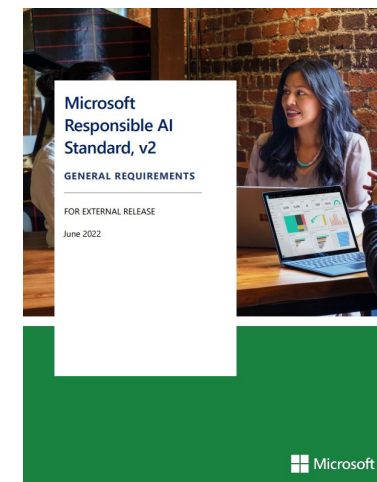
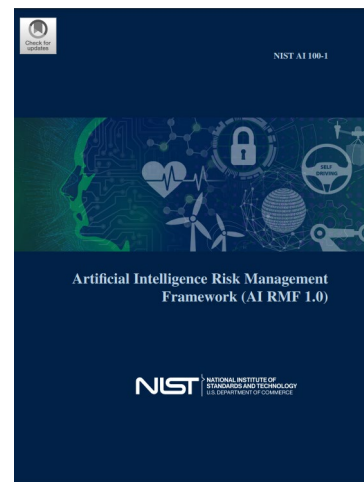


# Policy & Governance

## Introduction

While AI has existed for more than 50 years, policy and governance considerations for the secure use and design of AI is a relatively bleeding-edge topic for many practitioners. That said, security practitioners have long been deeply familiar with the overarching governance needed to help align security concerns with business requirements and customer expectations, as well as relevant laws and regulations. Better practices within traditional secure SDLC and model risk can serve as the foundations of more specific, AI-focused security practices. According to the OWASP AI Security & Privacy Guide, “AI Security does not mean abandonment of prior security practices – industry better practice is to continue existing security programs and to augment for nuances of AI.”<sup>1</sup>

For those considering how to embed AI Security principles across their businesses, it is also important to consider the existing industry frameworks and guidance that have been released over the last two years. These include NIST AI Risk Management 1.0, MITRE ATLAS™, Microsoft Responsible AI Standard v2, OWASP AI Security and Privacy Guide, ISO / IEC 23894:2023, ISO/IEC DIS 5338 (under development), and ENISA Cybersecurity of AI and Standardization. Regulations such as the EU AI Act or AI Bill of Rights will put even more of an emphasis on integrating AI Security into policy.





Home



Introduction



Related Summit Sessions



Contributors



User Guide



Securing AI



Risk & Compliance



Policy & Governance



AI Bill of Materials



Trust & Ethics

# Policy & Governance

## Theme #1: Defining the parameters of AI Security for your organization

When organizations begin to consider policy and governance as they relate to AI security, they may feel overwhelmed by the spectrum of competing priorities, which, despite the intensity of recent media attention, should extend way beyond workforce use of ChatGPT. Critical areas of concern may include determining appropriate usage policies for Large Language Models (LLMs), anticipating how adversaries could leverage AI to launch attacks, using AI-enhanced technologies to bolster their own security efforts, and more. For policy and governance efforts to be effective, organizations must align on the definitions of AI security as they apply to their organizations.

So how do we get to meaningful, structured conversation around the parameters of AI Security Policy and Governance?



## Recommendations

In framing these conversations, practitioners will benefit from defining scope upfront for their intended audience. Our workshops revealed four major areas of cross-over between AI and security that practitioners can refer to in their own conversations:

- **Cybersecurity of AI:** Robustness and vulnerabilities of AI models and algorithms (ENISA Cybersecurity of AI and Standardisation)<sup>2</sup>
  - Subset includes differentiation between the enterprise's models and algorithms versus those of third parties (software, data, or hardware) (NIST.AI.100-1, Appendix B, Page 36)<sup>3</sup>
- **AI-Enabled Cybersecurity:** Leveraging AI to further advance or provide future autonomous operation of existing security practices (ENISA Cybersecurity of AI and Standardisation)<sup>2</sup>
- **Adversarial AI:** Adversaries exploit vulnerabilities of AI systems to alter behavior to serve a malicious end goal (MITRE ATLAS)<sup>4</sup>
- **Malicious Use of AI:** Malicious use of AI to create more sophisticated attacks (ENISA Cybersecurity of AI and Standardisation)<sup>2</sup>





Home



Introduction



Related Summit Sessions



Contributors



User Guide



Securing AI



Risk & Compliance



Policy & Governance



AI Bill of Materials



Trust & Ethics

# Policy & Governance

## Theme #2: Determining ownership and contributions

While the challenges of ownership are not a new issue for security organizations, defining who is responsible for communicating to the business the recommended policy-level direction for AI Security requires a few unique considerations.



Since AI will often be utilized across many different business units, security practitioners should be key contributors to policy efforts, but ownership of the policies themselves should reside outside of the CISO organization. This organizational model aligns with the expectations of ethical use set forth by the EU AI Act and industry frameworks such as the Responsible AI Institute Implementation Framework.<sup>5</sup>



## Recommendations

As organizations approach AI Security policy and governance, they should leverage the following principles:

- **Obtain executive buy in and public support:** While many initial conversations around AI security may stem from news coverage and hype, formal policies for the organizational use of AI need to come from the highest levels of leadership.
- **Draw from broad “Responsible AI” principles:** Organizations would be well served modeling their internal AI security policies after broader “Responsible AI” principles. According to the NIST Artificial Intelligence Risk Management Framework (AI RMF 1.0)<sup>3</sup>, such principles comprise “technology that is...equitable and accountable,” a commitment to allowing future generations to meet their own needs, “organizational practices [that] are carried out in accord with “professional responsibility,” and an approach that “aims to ensure that professionals who design, develop, or deploy AI systems and applications or AI-based products or systems, recognize their unique position to exert influence on people, society, and the future of AI.”
- **Reflect multiple areas of expertise in policy drafting:** Policies that can be adopted across the organization will require integration into many workstreams, including Data Science, Security, Quality, Procurement, Risk, and Legal, as recommended by the NIST AI RMP referenced above and the ENISA<sup>2</sup> Cybersecurity of AI and Standardization 5.2.2. The NIST guidelines stress the importance of having general counsel draft internal policies and practices with input from the CISO organization.





Home



Introduction



Related Summit Sessions



Contributors



User Guide



Securing AI



Risk & Compliance



Policy & Governance



AI Bill of Materials



Trust & Ethics

# Policy & Governance

## Theme #3: Increasing organizational awareness when codifying policies

It is clear the potential security risks associated with leveraging AI are not yet common knowledge. Therefore, security teams should impress upon leadership the importance of educating business unit leaders, software engineers, data engineers, data scientists, procurement, legal counsel, and others.

- 
- 
- 
- 

Through this approach, there will be a baseline understanding of risks, a workforce that is empowered to comply with policies as they go about their day-to-day responsibilities, and champions for adoption throughout the organization long before formalized policies are in place.



## Recommendations

When preparing education and awareness materials, consider the following:

- **Define known AI Security risks:** Inform your perspective on the expanded attack surface associated with AI by taking inventory of known AI Security risks and tailoring your educational materials to your audiences' level of understanding. As outlined by MITRE ATLAS,<sup>6</sup> AI-associated risks include reconnaissance, ML Model Access, ML Attack Staging, Defense Evasion, and others. Other techniques, as detailed in the OWASP AI Security and Privacy Guide,<sup>1</sup> include data poisoning, input manipulation, model inversion and theft, membership inferences, and more.
- **Provide accessible awareness and risk management sessions:** As outlined in the NIST AI RMF,<sup>3</sup> organizations should provide opportunities for teams to gain a thorough understanding of what AI Security practices will be expected of them upon policy implementation.





Home



Introduction



Related Summit Sessions



Contributors



User guide



Securing AI



Risk & Compliance



Policy & Governance



AI Bill of Materials



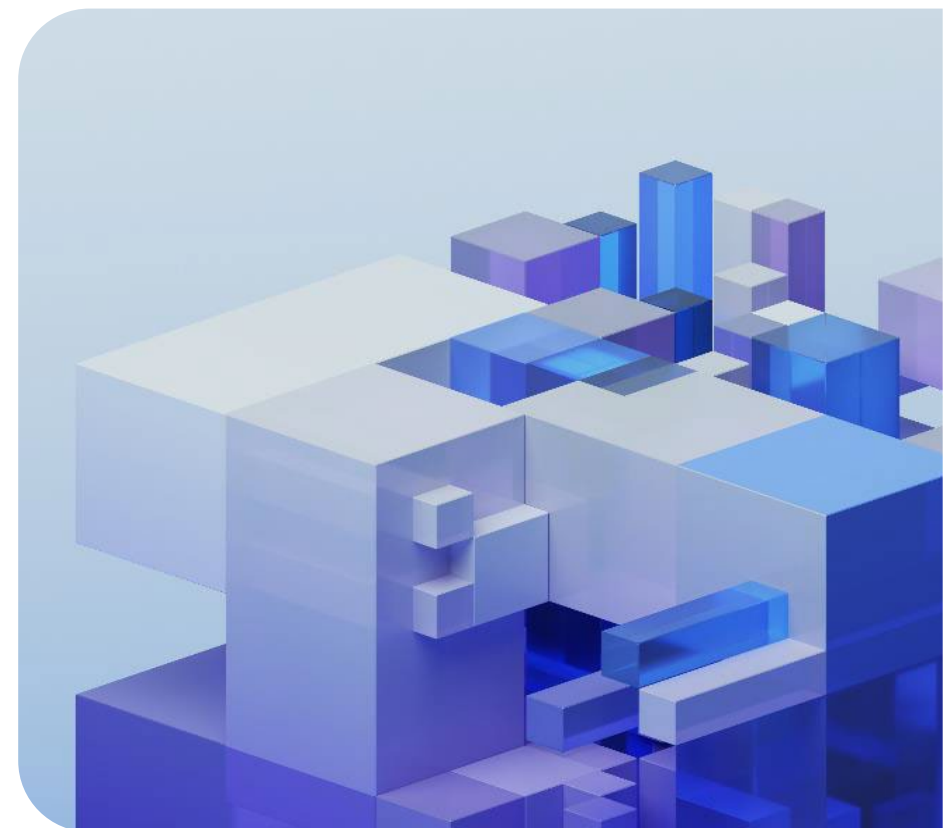
Trust & Ethics

# AI Bill of Materials

## Introduction

As emphasized in prior sections, it is imperative for AI Security practitioners to understand the dynamic nature of AI life cycle components and gain transparency into AI usage across their organizations. Thus, our next topic – the AI Bill of Materials (AI-BOM). In this guide, our reference to the AI Bill of Materials is focused on understanding all of the upstream components (i.e., training data, pre-trained models, vendor pedigree of those components, etc.) that informed the AI model. For example, some organizations are not developing their own AI algorithms but are instead leveraging pre-existing models and training that model on data sets they aggregated, data sets they purchased, or a combination of both.

Practitioners should also be challenging themselves to look at external AI-BOMs with which they may be interacting, so they have better visibility into their true risk exposure landscape. This could mean that organizations inquire into the AI-BOM for web-based LLMs that are using their data for training, even if the source code is private. Or it could mean asking for clarity on the AI-BOMs that third parties are using to provide products/services to clients. Finally, it is important to understand whether a model being leveraged was trained first by another organization and then trained by another in a sequential fashion so there is transparency into potential risks along the value chain.





Home



Introduction



Related Summit Sessions



Contributors



User guide



Securing AI



Risk & Compliance



Policy & Governance



AI Bill of Materials



Trust & Ethics

# AI Bill of Materials

## Theme #1: Increasing visibility into where AI resides within the organization

The consensus among security practitioners is that it is difficult to accurately identify the state of AI in their organizations. Having this visibility is critical to allowing an organization to align appropriate controls. Organizational issues driving this challenge include:

- **The nature of the technology is such that it may not be subject to centralized governance.** Most organizations keep records of which AI platforms they license. However, since data scientists may be siloed in individual business functions, not centrally governed, the totality of AI models may not be widely known or documented sufficiently. Additionally, low code / no code AI platforms may allow the business to embed models without the knowledge of the data science or governance teams.
- **Third parties may not be transparent about AI models they leverage to provide products and services to clients.** In the traditional third-party assessment process, there is not a consistent mechanism for third parties to disclose whether they leverage AI in the products / services they provide to clients.
- **Organizations may interact with models that are privately owned but will combine their training data with others, e.g., ChatGPT.** In many organizations, employees are starting to use generative models like ChatGPT without obtaining documented business approval or considering associated risks.



## Recommendations

Most security teams have received inquiries into how their teams are gaining visibility into business usage of tools like LLMs. However, recognizing the scope of AI is much broader, it is important to consider some pathways to enhance transparency in a sustainable fashion. Recommendations include:

- **Form a steering committee comprising cross-functional stakeholders (e.g., business, legal, privacy, security, user experience, IT).** The committee could oversee the documentation, review, and approval of AI use cases as part of its mandate. And, ideally, the committee will be sponsored by the Board of Directors.
- **Explore technical means to detect the use of AI in the enterprise.** Some organizations have blocked AI applications like ChatGPT to track whether employees are attempting to use them. Other technical detection techniques include automated discovery via specific connected platforms to compel business units to submit use cases for formal review and approval.
- **Include questions on vendor use of AI in third-party assessments.** Work with your third-party risk management, procurement, and legal teams to update your questionnaires to include questions on the use of AI in vendor products and services. Consider the degree to which the organization has exposure to enterprise-developed models, third-party models that may or may not be federated, third-party models that were then retrained by the client organization, and AI-based products and services from fourth parties that are used to support third parties.





Home



Introduction



Related Summit Sessions



Contributors



User guide



Securing AI



Risk & Compliance



Policy & Governance



AI Bill of Materials



Trust & Ethics

# AI Bill of Materials

## Theme #2: Focusing on actions to pursue today to improve the organization's AI posture

Once practitioners begin to uncover the many ways in which AI could be leveraged by their organizations (directly or indirectly), it is important to prioritize those actions that can be pursued in the near term. At the same time, AI Security should be brought to the table to draft directive controls that align with the current risk landscape.

This need for directive controls is emerging today as organizations experience an onslaught of inquiries related to how they are educating their employees on AI risks, how they are controlling their attack surface, and how to embed security controls without preventing the business from unlocking innovation opportunities or keeping pace with competitors.



## Recommendations

For organizations that are beginning their AI journeys, it is important to keep humans in the loop, as illustrated by the following guidance:

- **Implement directive controls, and training & awareness programs, to remind staff of their ongoing obligations.** Organizations should share with employees the potential risk exposures to the business by introducing AI. This applies whether AI is developed within the enterprise, outside the enterprise, or by third parties.
- **Update relevant policies including the Acceptable Use Policy (AUP).** Practitioners should work across multiple stakeholder groups to get their input on updates to the AUP or creation of an amendment specific to AI that requires employees to submit requests for approval of AI usage outside of the scope of the AUP.
- **Update third-party due diligence to understand how security safeguards are implemented by third parties.** In the event third parties declare usage of AI in providing services / products to your organization, consider updating third-party assessments to inquire about security safeguards third parties employ when leveraging AI in their delivery.





Home



Introduction



Related Summit Sessions



Contributors



User guide



Securing AI



Risk & Compliance



Policy & Governance



AI Bill of Materials



Trust & Ethics

# AI Bill of Materials

## Theme #3: Thinking two steps ahead

As organizations gain a deeper understanding of where AI is being leveraged beyond internal development, it is important to think about how to control the risks associated with using emerging technologies and solutions that allow secure innovation.

- 
- 
- 
- 

While in the short term, threat prevention and detection may rely more heavily on people, processes, and organizational policies, technical controls will allow organizations to evolve to incorporate better security practices and move away from heavy reliance on end users' judgement.



## Recommendations

When preparing to operationalize AI supply chain visibility, consider the following:

- **Leverage federated models for specific, sensitive use cases.** When training data is considered sensitive or confidential, it is wise to give serious consideration to using federated models. For example, life sciences companies may use federated models for clinical trials, where they need to have more control over the training data.
- **Consider potential use cases for leveraging firewall products.** Products that apply a firewall as a layer of protection between employee prompts and the generative AI model to better support policy enforcement and/or data security governance.
- **Explore privacy-preserving model learning.** To avoid the potential exposure of sensitive information (including PII, PHI, intellectual property, trade secrets, etc.), organizations will want to pursue AI training using encrypted data whenever possible. Practitioners should indicate in the AI Bill of Materials when and where they are leveraging anonymized data. And it is good practice to use anonymization techniques to ensure sensitive or confidential information is not retained within the black box.
- **Silo sensitive data where appropriate.** Consider data enclaves where personal data is stored in restricted secure environments.







Home



Introduction



Related Summit Sessions



Contributors



User Guide



Securing AI



Risk & Compliance



Policy & Governance



AI Bill of Materials



Trust & Ethics



# Trust & Ethics

## Introduction

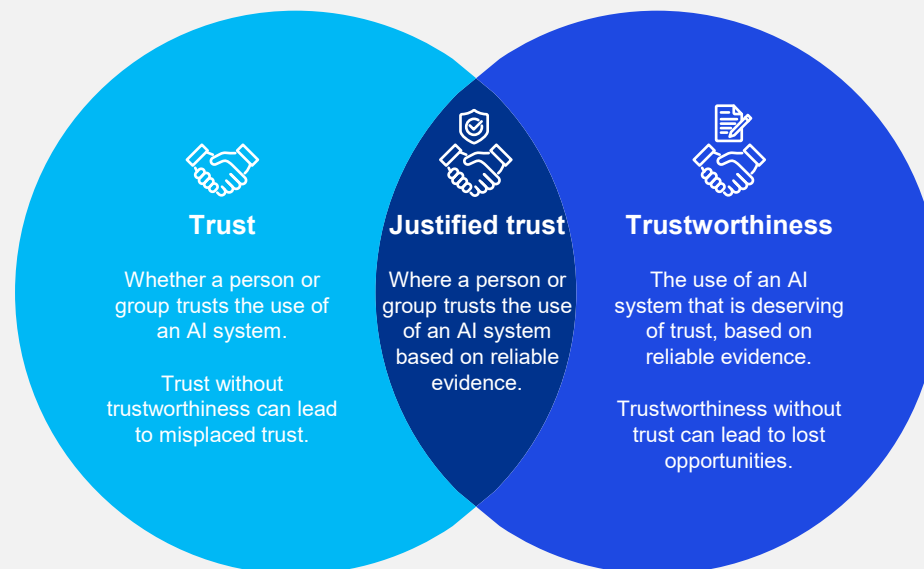
Trust and ethics are the foundations of cybersecurity professionals' responsibilities. ISC, a well-accepted industry provider of cyber certifications, has prioritized this in the first canon of their Code of Ethics<sup>8</sup>, "Protect society, the common good, necessary public trust and confidence, and the infrastructure."<sup>2</sup>

So how does this construct apply to the way security practitioners approach trust and ethics in the AI Security realm? Practitioners should consider their role in the broader framework of Responsible AI<sup>9</sup>, i.e., protecting and preserving the integrity of AI models and the privacy around training data and user input, which can comprise intellectual property and other sensitive data. Also critical is committing to transparency around AI processes and the AI Bill of Materials, which comprises data sources, models uses, and more. Such efforts will go a long way toward building trust around AI in general, bridging gaps between the known and unknown, and helping organizations navigate through uncertainty.

The centerpiece of AI trust and ethics is "justified trust" – which lies at the intersection of "trust" and "trustworthiness" and is captured in the graphic at the right from the Centre for Data Ethics and Innovation's AI Assurance Guide<sup>10</sup>.

## CDEI AI Assurance Guide – Justified Trust<sup>10</sup>

### The relationship between trust, trustworthiness, and justified trust





Home



Introduction



Related Summit Sessions



Contributors



User Guide



Securing AI



Risk & Compliance



Policy & Governance



AI Bill of Materials



Trust & Ethics



# Trust & Ethics

## Theme #1: Establishing the security team's role in AI trust & ethics

All organizations need to determine which areas of trust and ethics are the responsibility of the AI Security team versus other key stakeholders, such as the business or the data science team. Codifying where responsibility lies is critical, whether the decision is informed by industry-specific regulations, your organization's AI governance policies, or an early adopter approach associated with embedding security into responsible development and use of AI.



Per the discussion of "justified trust" on the previous page, security teams must understand which components they can contribute to, such as the principles of technical robustness and safety, as well as privacy and data governance within the European Commission's Ethics Guidelines for Trustworthy AI<sup>11</sup> or the dimensions of Data and System Operations, Consumer Protection, and Robustness within the RAIL's Certification Dimensions<sup>11</sup>.

These imperatives are supported by the following statement from the NIST AI RMF Risks and Trustworthiness 3.0<sup>3</sup>: *"It is the joint responsibility of all AI actors to determine whether AI technology is an appropriate or necessary tool for a given context or purpose, and how to use it responsibly. The decision to commission or deploy an AI system should be based on a contextual assessment of trustworthiness characteristics and the relative risks, impacts, costs, and benefits, and informed by a broad set of interested parties."*



## Recommendations

Delineation of roles and responsibilities in the pursuit of trusted, ethical AI systems should align with the subject matter expertise of the professionals involved.

- **Remember that privacy and security are components of, but not the entirety of, responsible / ethical AI.** Microsoft's Responsible AI Standard<sup>12</sup> is a notable example of ethics guidelines and responsible AI dimensions. The standard outlines goals for accountability, transparency, fairness, reliability and safety, privacy and security, and inclusiveness. When it comes to privacy, confidentiality and integrity, remember that these ideals should apply to not only the AI model itself, but also prompt data, training data, and model output.
- **Align with the business on their responsibility for the data ingested by AI, as well as the definition of expected outcomes.** The security team is not responsible for deciding what data is put into the model, as they will not have the same level of context that the business has around the intended purpose of the model.
- **Designate the security team as responsible for advising the business on potential threats to the confidentiality and integrity of data.** Although members of the security team may not have context for the data itself, they can still serve as advisors on integrity. For example, they will be aware of whether data sits in a data lake with substandard access controls and can advise whether data has been classified as restricted and may, therefore, require different handling or be excluded from usage in the model.





Home



Introduction



Related Summit Sessions



Contributors



User Guide



Securing AI



Risk & Compliance



Policy & Governance



AI Bill of Materials



Trust & Ethics



# Trust & Ethics

## Theme #2: Supporting the CIA triad continuously and consistently

To foster ethical and trustworthy AI, security professionals should continue to support the CIA (Confidentiality, Integrity, Availability) triad as they apply to AI models, as well as to prompt data, training data, and model outputs. Each of the latter may contain sensitive information that should be protected from tampering and exposure.

- 
- 
- 
- 

While security teams don't typically weigh in on the demographic inclusiveness of data sets or whether models are biased, it is appropriate for them to be involved in efforts to ensure that AI technologies align with standards of trustworthiness and ethics.



## Recommendations

To support trust & ethics through consistent commitment to the CIA triad, consider the following:

- **Apply data governance and controls continuously.** Traditional data governance principles apply to data used in training. Security professionals should consider the following guidance from the AWS Well-Architected Framework<sup>13</sup>:
  - Identify and classify sensitive data
  - Tag resources and models made from sensitive elements
  - Encrypt sensitive data
  - Anonymize or de-identify data where possible
- **Limit access to training datasets.** Determine which individuals should have access to or interact with training data sets or the AI model itself. Consider the following from the NIST AI RMF AI Risks and Trustworthiness 3.3 (NIST.AI.100-1)<sup>3</sup>:
  - Limit access to data to engineers that need access to data, or that leverage AI platforms that do not give data scientists access to the data (OWASP AI Security and Privacy Guide)<sup>1</sup>





Home



Introduction



Related Summit Sessions



Contributors



User Guide



Securing AI



Risk & Compliance



Policy & Governance



AI Bill of Materials



Trust & Ethics



# Trust & Ethics

## Theme #3 : Considering security's support of privacy considerations

A key underlying principle of securing AI models is that using personal data for AI training should be conducted fairly and lawfully, per both U.S. and EU regulatory guidelines.

- Organizations should consider security's role in ensuring privacy issues such as this are addressed in the enterprise approach to ethical and responsible AI development. Further, although the EU AI Act is still undergoing edits and formal voting at the time of this publication, regulations like GDPR are raising questions among security professionals about how privacy principles will be applied to the black box of AI. With transparency challenges, there is a need for more assurance that there are proper controls in place, ongoing monitoring, and periodic evidence to maintain trust.

There could be negative consequences for organizations that leverage data outside of the original purpose communicated to individuals sharing their data. Key privacy and data governance resources include the European Commissions' Ethics Guidelines for Trustworthy Artificial Intelligence; guidance from the Responsible AI Institute on consumer protection; and the NYU Center for Responsible AI, which provides awareness / education, frameworks, data considerations, tools, and more.



## Recommendations

The confluence of privacy regulations, unique nuances of the AI lifecycle, and responsible AI guidance will create complexities for security and privacy professionals. Until regulators provide formal statements on how privacy regulations apply to AI, teams should consider taking the following actions:

- **Begin discussions with your AI, privacy, legal, and security teams.** The privacy and legal teams should provide transparency into which regulations and legal obligations the security organization will be responsible for. AI subject matter experts can provide line of sight into the nuances of your organization's AI lifecycle development. Finally, security can provide the controls to allow for "security by design" while also identifying gaps. For example, previously established retention standards may need to be revisited to allow AI models to provide desired insights.
- **Begin to model the potential business threats that some AI privacy principles may introduce to your organization.** Ensure your team models potential risks around application of privacy principles, such as unlearning<sup>14</sup> upon request associated with the "right to be forgotten" principle or the rectification of an individual's record. In this example, an attacker can leverage unlearning to find weaknesses to infer information about the individual whose data was revoked from the training data set, as well as the privacy degradation as a result of unlearning.





Home



Introduction



Related Summit Sessions



Contributors



User guide



Securing AI



Risk & Compliance



Policy & Governance



AI Bill of Materials



Trust & Ethics

# Practitioner Guide Resources

#	Resource
1	OWASP; AI Security & Privacy Guide, (2023).
2	European Union Agency for Cybersecurity (ENISA); Cybersecurity AI and Standardization, (2023).
3	National Institute of Standards and Technology(NIST); Artificial Intelligence Risk Management Framework(AI RMF 1.0), (2023).
4	MITRE ATLAS; Adversarial Machine Learning 101, (2023).
5	AI Responsible Artificial Intelligence Institute; AI vs. Responsible AI: Why is it Important?, (2023).
6	MITRE ATLAS; Tactics, (2023).
7	European Parliament; EU AI ACT Draft Report, (May 2023).
8	(ISC)2; What Do You Do When No One is Watching, (2023).
9	AI Responsible Intelligence Institute; Accelerating Your AI Journey, (2023).
10	Centre for Data Ethics & Innovation; The Need for Trust in AI Systems, (2023).
11	European Commission; Ethics Guidelines for Trustworthy AI, (April 2019).
12	Microsoft; Responsible AI Standard v2, (June 2022).
13	AWS; MLSEC-04: Protect Sensitive Data Privacy, (2023).
14	Cornell University; When Machine Learning Jeopardizes Privacy, (September 2021).
15	AI & Data Today; What is the Cognitive Project Management for AI (CPMAI) Methodology?, 2023.





Home



Introduction



Related Summit Sessions



Contributors



User guide



Securing AI



Risk & Compliance



Policy & Governance



AI Bill of Materials



Trust & Ethics

# Contact Us



**Katie Boswell**  
Managing Director  
*AI Security Lead*  
T: +1 908 433 3417  
E: [katieboswell@kpmg.com](mailto:katieboswell@kpmg.com)



**Matt Miller**  
Principal  
T: +1 571 225 7842  
E: [matthewpmiller@kpmg.com](mailto:matthewpmiller@kpmg.com)



**Kristy Hornland**  
Director  
T: +1 425 281 5251  
E: [khornland@kpmg.com](mailto:khornland@kpmg.com)



**Mark Orsi**  
CEO  
T: +1 973 879 1200  
E: [morsi@grf.org](mailto:morsi@grf.org)



**Jonathan Dambrot**  
CEO  
T: +1 908 361 6438  
E: [jdambrot@cranium.ai](mailto:jdambrot@cranium.ai)





The KPMG name and logo are trademarks used under license by the independent member firms of the KPMG global organization.



[kpmg.com/socialmedia](https://kpmg.com/socialmedia)

The information contained herein is of a general nature and is not intended to address the circumstances of any particular individual or entity. Although we endeavor to provide accurate and timely information, there can be no guarantee that such information is accurate as of the date it is received or that it will continue to be accurate in the future. No one should act upon such information without appropriate professional advice after a thorough examination of the particular situation.

© 2023 KPMG LLP, a Delaware limited liability partnership and a member firm of the KPMG global organization of independent member firms affiliated with KPMG International Limited, a private English company limited by guarantee. All rights reserved. USCS000515-1A

The KPMG name and logo are trademarks used under license by the independent member firms of the KPMG global organization.