MITRATECH

WHITEPAPER

Navigating Global AI Regulations: Implications and Advice for Third-Party Risk Management Programs

Table of Contents

Introduction

Over the last three years, the world has witnessed a surge in the integration of artificial intelligence (AI) technology into existing solutions and new platforms, making AI a central component of their value proposition. Regulatory regimes spent years catching up to these new technologies, which also included proponents of the technology praising its transformative potential and detractors advocating for caution in how AI was used, considering the risks it posed.

The discussion about how to responsibly develop AI-based technologies has dominated discourse for years now. There are risks with unrestricted AI technology across several dimensions, including:

- **Data privacy concerns** – Sensitive data may be unintentionally exposed when it comes to training datasets.
- **AI governance concerns** – There is limited visibility into how AI is utilized in organizations and how it can be used responsibly.
- **Bias in AI outputs** – Generative AI uses extensive amounts of data to train its models and create results. If the data fed into the model is overly biased in any direction, that can also result in unintentional bias in the outputs.
- **Transparency in model development and outputs** – AI developers often fail to provide sufficient insight into how their models arrive at their decisions. This lack of visibility into algorithm operation is an area of concern.

The pace of regulatory action on AI has started to rise around the world. Regulatory bodies in the United States, the United Kingdom, the European Union, and Canada have either proposed or finalized documents emphasizing outright restrictions, conscientious development, and approval processes. By 2028, more than 50% of developed countries are expected to have enacted regulations governing generative AI, Gartner research estimates. The different focuses across multiple geographies create a patchwork of regulatory complexity for third-party risk managers and cybersecurity professionals seeking to strike a balance between efficiency gains and responsible development.

This white paper examines key AI regulations and recommendations worldwide, providing guidance on integrating AI risk management into your broader third-party risk management program.

Understanding Key Terms

Although AI is not a new term, increasing awareness is largely due to the release of OpenAI's ChatGPT, which has introduced new terminology into our collective lexicon. Let's start with the basics.

- **Artificial intelligence (AI)** is a field of computer science that focuses on creating machines capable of performing tasks typically requiring human intelligence. AI systems use algorithms and large datasets to process information, recognize patterns, and make informed predictions or recommendations. Examples of AI include manufacturing robots, self-driving cars, and virtual travel booking agents.

- **Machine learning (ML)** is a term often interchanged with AI. However, ML is a subset of AI that enables computers to learn from data and improve their performance over time without being explicitly programmed. It's like teaching a computer to recognize patterns and make predictions based on examples it has seen before. ML examples include Google Translate, business intelligence and data analytics software, chatbots, and your Netflix recommended list.

- **Generative AI** is a type of artificial intelligence technology that describes machine learning systems capable of generating text, images, code, or other types of content, often in response to a prompt entered by a user. Generative AI applications are built on **large language models (LLMs)**, deep learning algorithms that can perform a variety of tasks using natural language queries, called **natural language processing (NLP)**. Large language models are trained using massive datasets — hence the name. This enables them to recognize, translate, predict, or generate text or other content. Generative AI models are increasingly being incorporated into online tools and chatbots that enable users to type questions or instructions into an input field, with the AI model generating a human-like response.

  Examples include ChatGPT, Claude, Google Gemini, DALL-E, and others.

Overview of Artificial Intelligence Regulations Worldwide

The recently released artificial intelligence regulations and guidance have so far primarily emphasized the responsible development of associated technologies. This has included addressing the risks presented by existing data privacy rules related to AI applications, as well as governing which use cases software companies are legally permitted to pursue without special government dispensation.
This also has implications for third-party risk managers because they need to consider which questions to ask vendors and suppliers about their use of artificial intelligence or machine learning algorithms within their own business operations as part of vendor assessments.

The rest of this section will examine the following regulations and guidelines in the context of TPRM:

- The European Union Artificial Intelligence Act
- National Institute of Standards and Technology (NIST) guidance documents on AI
- ISO 42001 (AIMS Framework)
- The United Kingdom's Artificial Intelligence Regulation Bill
- The Artificial Intelligence and Data Act (Canada)

Each of these regulations and standards will be covered in its own section, and its implications will be examined separately for businesses covered by the regulation.

The European Union Artificial Intelligence Act

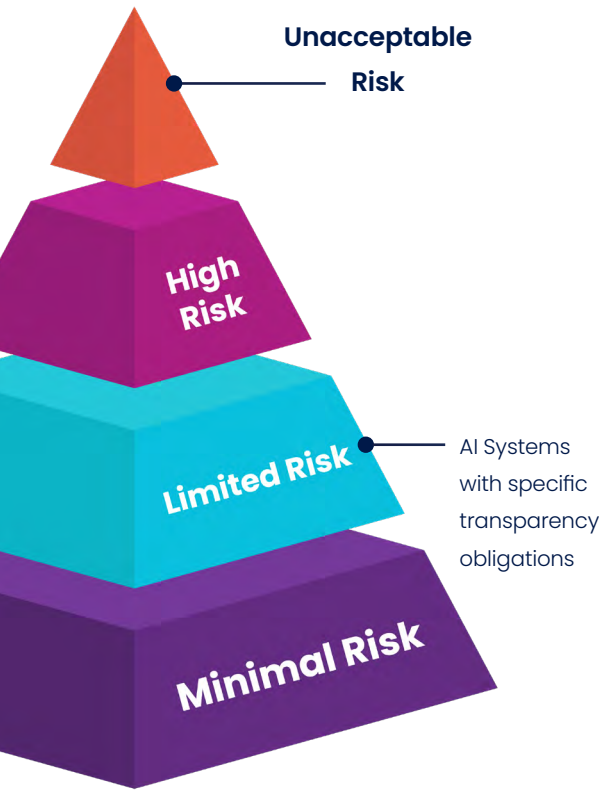
The EU approved the first official AI regulations in the world in July 2024, marking a historic moment in AI legislation. This is not the first time that Europe has preceded other regulatory regimes. In 2016, the EU became the first region to implement comprehensive data privacy rules with the introduction of the General Data Protection Regulation (GDPR). The AI Act provides comprehensive legislative guidance on the responsible development of AI, outlining which applications can and cannot be developed without special permission, as well as how those applications can be utilized.

The European Union's Artificial Intelligence Act, officially the "Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts," was originally proposed in 2021. It passed in June 2024, and initial requirements went into effect in February 2025, with additional measures going into effect later in 2025 and 2026.

The rules in the legislation are designed to:

- Address risks specifically created by AI applications.
- Propose a list of high-risk applications.
- Set clear requirements for AI systems for high-risk applications.
- Define specific obligations for AI users and providers of high-risk applications.
- Propose a conformity assessment before the AI system is put into service or placed on the market.
- Propose enforcement after such an AI system is placed in the market.
- The European Parliament adopted a risk-based approach to the regulation.

The European Parliament adopted a risk-based approach to the regulation. The rules define four categories of risk placed in a pyramid.



These four levels are defined as:

Unacceptable Risk – This level refers to any AI systems that the EU considers a clear threat to the safety, livelihoods, and rights of EU citizens. Two examples are social scoring by governments, which appears to ban one of the common uses of AI in China, and toys that use voice assistance to encourage dangerous behavior.

High Risk – AI systems marked as high risk serve functions viewed as critical to society. This can include AI used in education, employment practices, law enforcement, border control and immigration, critical infrastructure, and other situations where someone's rights might be infringed.

Limited Risk – This category refers to AI applications with specific transparency obligations, such as a chatbot on a website that must inform users they're interacting with a software program. The goal is to empower consumers and others to decide whether or not to interact with the AI system.

Minimal Risk – Also called "no risk," these are AI systems used in media like video games or AI-enabled email spam filters. This appears to be the bulk of AI used within the EU today.

High-risk AI systems have specific, strict rules that they must comply with before they can go on the market. According to the EU's write-up on the AI Act, high-risk systems must include the following:

- Adequate risk assessment and mitigation systems.
- High-quality datasets feeding the system to minimize risks and discriminatory outcomes.
- Logging of activity to ensure traceability of results.
- Detailed documentation providing all information necessary on the system and its purpose for authorities to assess its compliance.
- Clear and adequate information to the user.
- Appropriate human oversight measures to minimize risk.
- High level of robustness, security, and accuracy.

All remote biometric systems, for example, are considered high risk. The use of remote biometrics in public spaces for identification (e.g., facial recognition) in law enforcement will be prohibited under this act.

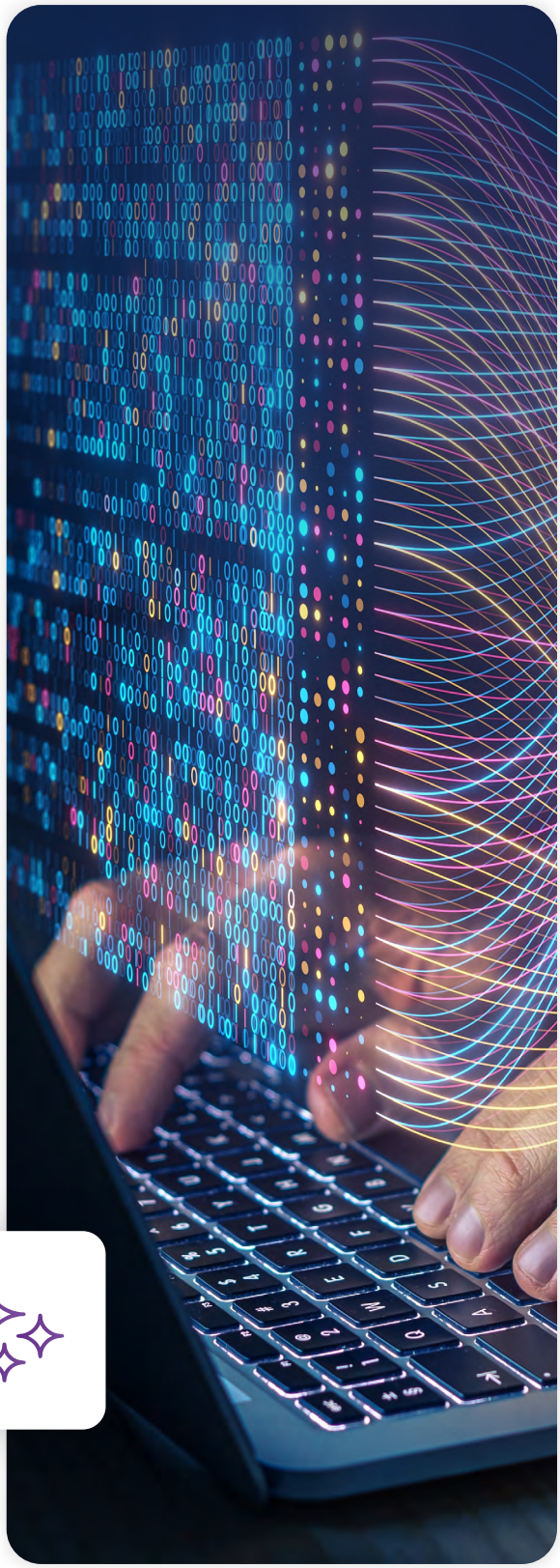
The AI Act establishes a legal framework for reviewing and approving high-risk AI applications, aiming to protect citizens' rights, minimize bias in algorithms, and control negative AI impacts.

Impact on Third-Party Risk Management Programs

Companies located in the EU or those with EU customers and third-party partners must carefully study the new legislation to ensure compliance with the included standards. As part of formalizing the rule, the EU created a website that features the complete text of the law and a basic compliance checker to see how the AI Act affects any specific AI system.

The expansive definition of "high risk" means that companies need to expand their vendor risk assessment questionnaires to understand how well vendors and suppliers comply with the new rules. This ruling could alter how AI is deployed in the EU and what applications ultimately get developed or not developed. Organizations should also thoroughly examine their own AI implementation practices. Other European technology laws still apply, so companies that need to comply with GDPR should also explore ways to integrate AI Act compliance into their workflow.

Each of these regulations will be covered in its own section, and its implications will be examined separately for businesses covered by the regulation.



Artificial Intelligence Goverance in the United States

Currently, there is no federal legislation governing the development of AI systems in the United States. Each administration has issued various executive orders regarding various facets of AI usage, development, and deployment, but enforceable regulations have been left to individual states. In the 2025 legislative session, according to the National Conference of State Legislatures, AI bills were introduced or considered in all 50 states, Puerto Rico, the U.S. Virgin Islands, and Washington, D.C.; 28 states and territories enacted legislation or adopted resolutions. Some states, like Colorado, enacted legislation requiring limitations on potential algorithmic bias, while others, like Florida, provided grants to school systems seeking to implement AI tools in their operations.

The lack of federal regulations does not mean there is no activity. Federal standards bodies have issued extensive guidance. For instance, the National Institute of Standards and Technology (NIST) has published six guidance documents on artificial intelligence and one open-source tool (Dioptra) designed to check claims about AI model operation. These documents are:

- The AI Risk Management Framework (AI RMF)
- Preventing Misuse of Dual-Use Foundation Models (NIST AI 800-1)
- Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile (NIST AI 600-1)
- Secure Software Development Practices for Generative AI and Dual-Use Foundation Models (NIST Special Publication (SP) 800-218A)
- A Plan for Global Engagement on AI Standards (NIST AI 100-5)

The documents released from the National Institue of Standards and Technology (NIST) are tailored specifically to mitigate the risks of artificial intelligence rather than provide hard and fast rules about how companies can leverage the technology, whether internally or in products, to sell. Regardless of the focus, these guidance documents are vital for third-party risk managers to review and apply to their vendor landscape.

NIST AI Guidance Documents and Third-Party Risk Management

The National Institute of Standards and Technology is the leading technology guidance body in the United States. They have developed several guidance documents related to artificial intelligence, starting with the AI RMF in January 2023 as the core document and then five ancillary documents designed to support how companies think about the inherent risk of artificial intelligence technologies.

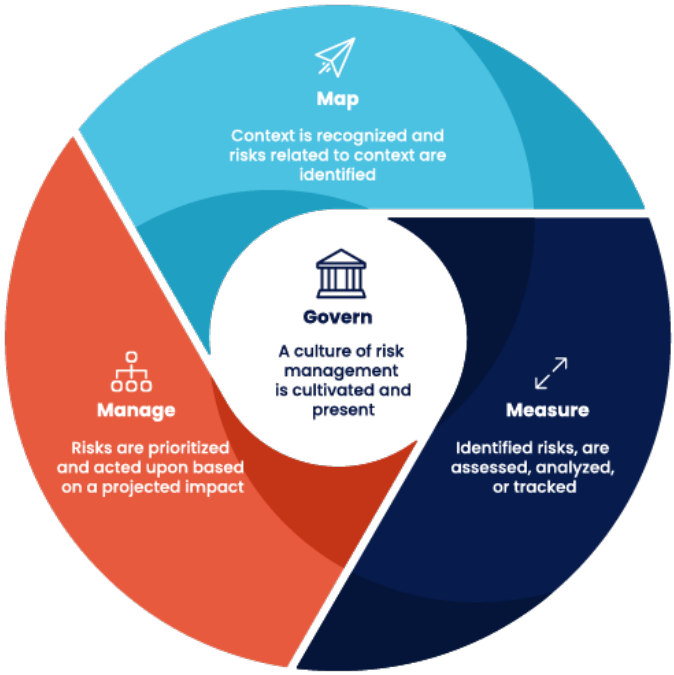
The NIST AI Risk Management Framework

Congress directed NIST to develop guidance around how to govern the risks associated with artificial intelligence. The outcome of that was the NIST AI Risk Management Framework, released in January 2023.

The framework offers guidance for organizations as they develop an internal AI governance strategy. This policy document is designed to support the use of AI throughout the organization, including within the third-party risk management ecosystem. The RMF is divided into two parts:

- **Part 1:** Foundational Information includes an overview of risks and characteristics of what NIST refers to as "trustworthy AI systems."
- **Part 2:** Core & Profiles describes four functions to help organizations address the risks of AI systems: Govern, Map, Measure, and Manage.

The illustration below reviews the four functions.



Organizations should consider risk management principles to minimize the potential negative impacts of AI systems, including hallucination, data privacy breaches, and threats to civil rights. This consideration also extends to the use of third-party AI systems or the use of AI systems by third parties. Potential risks of third-party misuse of AI include:

- Security vulnerabilities in the AI application itself. Without the proper governance and safeguards in place, your organization could be exposed to system or data compromise.
- Lack of transparency in methodologies or measurements of AI risk. Deficiencies in measurement and reporting could lead to an underestimation of the impact of potential AI risks.
- AI security policies that are inconsistent with other existing risk management procedures. Inconsistency results in complicated and time-intensive audits that could introduce potential negative legal or compliance outcomes.

According to NIST, the RMF will help organizations overcome these potential risks. Throughout the document, NIST outlines specific actions to take in each phase, as well as potential risks that could arise from the use of artificial intelligence. The AI RMF serves as the core document for NIST's AI risk mitigation strategy.

Impact on Third-Party Risk Management

The NIST AI RMF is designed to help organizations identify the risks associated with AI usage in their internal operations and supply chain. It's vital for companies to consider these risk management principles to minimize the potential negative impacts of AI systems, such as hallucination, data privacy, and threats to civil rights. This consideration also extends to the use of third-party AI systems or third parties' use of AI systems. Potential risks of third-party misuse of AI include:

- Security vulnerabilities in the AI application itself. Without the proper governance and safeguards, your organization could be exposed to system or data compromise.
- Lack of transparency in methodologies or measurements of AI risk. Deficiencies in measurement and reporting could result in underestimating the impact of potential AI risks.
- AI security policies are inconsistent with other existing risk management procedures. Inconsistency results in complicated and time-intensive audits, which could introduce potential negative legal or compliance outcomes.

The AI RMF serves as a baseline for resolving these issues, particularly in light of the additional guidance released by NIST for contextualizing the risks associated with artificial intelligence.

NIST AI RMF Generative Artificial Intelligence Profile (NIST AI 600-1)

The Generative AI profile publication aims to highlight the cross-sector risks explicitly associated with generative AI.

The profile is broadly applicable to all industries, which is why it's considered a cross-sectoral profile within NIST. According to NIST, cross-sectoral profiles can be used to govern, map, measure, and manage risks associated with activities or business processes common across sectors. In the case of Generative AI, these risks can occur with the use of large language models (LLMs), cloud-based services, or acquisitions. The profile outlines a non-exhaustive list of 12 risks associated with generative artificial intelligence that companies should be aware of when utilizing such models. These include:

- Hallucinations that occur from confidently stating incorrect information.
- Easier access to or generation of potentially dangerous information, such as data about weapons, disinformation, or hateful content.
- Data privacy risks from potential leakage or loss of sensitive information.
- Environmental risks from high compute allocations to use the models.
- Lowered barriers for offensive cyberattack potential, including automated discovery of vulnerabilities and exploit creation.

The GenAI profile also provides guidance based on the AI RMF for managing and mitigating risks related to generative AI. Much of this relates back to the Govern phase of the AI RMF, offering guidance on how to craft policies for managing the risks of generative AI and ensuring that GenAI tools are used in a transparent and effective manner internally.

What Does this Mean for TPRM?

Most relevant for third-party risk managers is the risk to value chain and component integration called out in the profile. This relates to the non-transparent or untraceable integration of upstream third-party components. Often, the biggest risk in the software supply chain is where the software components come from; the Log4j vulnerability of December 2021 demonstrated how damaging vulnerable open-source packages could be to companies. This risk is increased for GAI because it may generate components where tracing the provenance is functionally impossible.

A Plan for Global Engagement on AI Standards (NIST AI 100-5)

Another component of these guidelines is creating a framework for U.S. government collaboration with other standards bodies on artificial intelligence standards. NIST AI 100-5 outlines NIST's approach to collaborating with regulatory bodies and private organizations worldwide to develop unified regulatory standards for AI.

What Does this Mean for TPRM?

For third-party risk managers, this could lead to a harmonization in AI standards globally. In AI 100-5, NIST notes that the U.S. federal government is one of many stakeholders looking at implementing regulations around artificial intelligence development and safeguards against misuse. The goal of this publication is to put guardrails around some sense of international consistency in regulations.

Should this consistency pan out, it could simplify third-party risk analysis as it pertains to AI. A global harmonization of regulations means that compliance with regulations becomes easier across jurisdictions, making it simpler to integrate questions into vendor risk assessments and judge multinational compliance.

NIST AI Guidelines for Software Development

Two of the five NIST publications released since the AI Risk Management Framework were published focus on guidance for developers creating AI applications. Preventing Misuse of Dual-Use Foundation Models (NIST AI 800-1) and Secure Software Development Practices for Generative AI and Dual-Use Foundation Models (NIST Special Publication (SP) 800-218A) provide best practices for developers and foundation model creators to create more secure software with AI included.

What Does this Mean for TPRM?

From a third-party risk management perspective, these guidelines serve as information to understand what sorts of questions to ask vendors and suppliers about the usage of AI in their operations. NIST AI 800-1 especially covers misuse risks from a technical as well as societal perspective, providing guidance on how to add safeguards into foundation models to ensure they can't be used for predictable malicious purposes.

NIST SP 800-218A is an addendum to the Secure Software Development Framework specific to generative AI and foundation models. Its relevance to TPRM becomes apparent when asking vendors about their software development practices. It's also important to use when understanding how vendors have secured their AI models and access to data.

State-Based AI Governance

In lieu of federal legislation, many individual states are implementing regulations around artificial intelligence. Other states, such as Massachusetts, have clarified how their existing laws apply to developers, suppliers, and users of artificial intelligence. Recently enacted AI governance state laws include Utah's AI Policy Act, which went into effect May 2024, Colorado's AI Act (effective February 2026), and California laws AB 2013 and SB 942 (effective January 2026). These laws mandate greater transparency in the development and use of AI systems, require clear consumer disclosures, and explicitly prohibit algorithmic discrimination.

A few common themes include:

- **Transparency & Clear Disclosures:** Verify that our vendors clearly disclose AI use and are upfront with consumers about significant AI-driven decisions.
- **Fairness & Non-Discrimination:** Assess whether vendors regularly test for and address algorithmic biases.
- **Documentation & Impact Assessments:** Require detailed compliance documentation, including training datasets, impact assessments, and consumer protection practices.
- **Data Protection:** Emphasize security controls vendors must implement to protect personal information.
- **Regulatory Alignment:** Encourage vendors to align with recognized standards (like the NIST AI Risk Management Framework) to support compliance.

Current State Regulations May Apply to Artificial Intelligence

On April 16, 2024, the Massachusetts Attorney General clarified that existing consumer protection laws apply to AI. The guidance warns against making false claims about the capabilities, reliability, or safety of AI systems and reinforces compliance with privacy and anti-discrimination laws. It also highlights disclosure requirements under the Equal Credit Opportunity Act when AI is used in credit decisions. Third-party risk managers should expect similar enforcement trends across other states and ensure vendors' AI tools meet legal and ethical standards.

Risk management teams should verify that third-party vendors clearly disclose when customers interact with AI, regularly test their systems for biases, and comply with strict consumer protection and data security rules. Leveraging standards like the NIST AI Risk Management Framework can help mitigate liability risks under these new regulations.

U.S. Guidance Impact on Third-Party Risk Management

The United States' approach to AI has so far emphasized best practices guidance instead of direct regulation, leaving legislation up to individual states. According to Gartner research, it is estimated that **by 2030, 50% of the U.S. population will be covered under state-level AI regulations** in some capacity. This contrasts with the EU method of providing strict regulatory oversight for the development of AI technologies and applications more broadly.

The extensive collection of guidance documents developed in recent years offers many pathways forward for third-party risk managers. The Secure Software Development Framework (SSDF) extensions provide background for crafting questions related to development practices, while the AI RMF offers a framework to build upon. The ancillary guidance from NIST also ensures an extensive library of information on integrating AI risks into a TPRM program.

The United Kingdom's Artificial Intelligence Regulation Bill

In the United Kingdom, Lord Holmes of Richmond introduced an AI Regulation Bill in the House of Lords. This is the second bill introduced in Parliament designed to regulate the usage of artificial intelligence in the UK. The initial bill, addressing both AI and workers' rights, was presented in the House of Commons towards the conclusion of the 2022 to 2023 legislative session but was discontinued in May 2023 due to the session's conclusion.

The latest AI bill, introduced in November 2023, is broader in focus. Lord Holmes introduced the regulation to put some guardrails around AI development and define who would be responsible for defining future legislative restrictions on AI in the United Kingdom.

There are a few key features of the bill. These include:

- **The creation of an AI Authority** tasked with the primary responsibility of ensuring a cohesive approach to artificial intelligence across the UK government. It's also in charge of taking a forward-looking approach to AI and ensuring that any future regulatory framework aligns with international frameworks.
- **The definition of key regulatory principles** puts guardrails around the regulations that the proposed AI Authority can and should create. According to the bill, any AI regulations put in place must adhere to the principles of transparency, accountability, governance, safety, security, fairness, and contestability. The bill also notes that AI applications should comply with equalities legislation, be inclusive by design, and meet the needs of "lower socio-economic classes, older people, and disabled people."
- **Defining the need to establish regulatory sandboxes** that enable regulators and businesses to work together for effectively testing new applications of artificial intelligence. These "sandboxes" may also offer companies a way to understand and pinpoint appropriate consumer safeguards to do business in the UK.
- **Advocating for AI Responsible Officers in each company** seeking to do business in the UK. The role of this officer is to ensure that any applications of AI in the company are as unbiased and ethical as possible, while also ensuring that the data used in AI remains unbiased.
- **Guidelines for transparency, IP obligations, and labeling** stipulate that companies utilizing AI provide a record of all third-party data used to train their AI model, comply with all relevant IP and copyright laws, and clearly label their software as using AI. This component also grants consumers the right to opt out of having their data used in training AI models.

This proposed regulation is still in the early phases of negotiations. It could take a very different form after the second reading in the House of Lords, followed by a subsequent reading in the House of Commons.

Depending on how much of the bill survives the legislative process, it could have a substantial impact on how AI is used in the UK, how models are trained, and the transparency of the broader data-gathering process. Each of these areas has a direct impact on the usage of AI technologies by third-party vendors or suppliers.

The Artificial Intelligence and Data Act (Canada)

In June 2022, the government of Canada began consideration of the Artificial Intelligence and Data Act (AIDA) as part of Bill C-27, the Digital Charter Implementation Act, 2022. The larger C-27 bill is designed to modernize existing privacy and digital law and includes three different sub-acts: the Consumer Privacy Protection Act, the Artificial Intelligence and Data Act, and the Personal Information and Data Protection Tribunal Act.

The AIDA's main goal is to add consistency to AI regulations throughout Canada. There are a few regulatory gaps identified in the companion document of the Act, such as:

- Mechanisms such as human rights commissions provide for redress in cases of discrimination; however, individuals subject to AI bias may never be aware that it has occurred;
- Given the wide range of uses of AI systems throughout the economy, many sensitive use cases do not fall under existing sectoral regulators, and
- There is a need for minimum standards as well as greater coordination and expertise to ensure consistent protections for Canadians across use contexts.

The act is currently under discussion, and the Canadian government anticipates it will take approximately two years for the law to pass and be implemented. There were six core principles identified in the companion document to AIDA, outlined on the next page.

Guiding Principle	How AIDA Describes It	What It Could Mean for TPRM
Human Oversight & Monitoring	Human Oversight means that high-impact AI systems must be designed and developed to enable people managing the operations of the system to exercise meaningful oversight. This includes a level of interpretability appropriate to the context. Monitoring, through measurement and assessment of high-impact AI systems and their output, is critical in supporting effective human oversight.	Vendors and suppliers must establish easily measurable methods to monitor AI usage in their products and workflows. Following potential passage of the AIDA, organizations will need to understand how their third parties monitor AI usage and incorporate AI into their broader governance and oversight policies. Another way to ensure more thorough human oversight and monitoring is to build human reviews into reporting workflows to check for accuracy and bias.
Transparency	Transparency means providing the public with appropriate information about how high-impact AI systems are being used. The information provided should be sufficient to allow the public to understand the capabilities, limitations, and potential impacts of the systems.	Organizations should be asking their vendors and suppliers about how they're using AI and what sort of data is included in the models. Be aware of how this is integrated as well.
Fairness and Equity	Fairness and Equity means building high-impact AI systems with an awareness of the potential for discriminatory outcomes. Appropriate actions must be taken to mitigate discriminatory outcomes for individuals and groups.	Organizations should inquire how their third parties are controlling for potential bias in their AI usage. There might be an additional impact here in terms of net new ESG regulations.

Guiding Principle	How AIDA Describes It	What It Could Mean for TPRM
Safety	Safety means that high-impact AI systems must be proactively assessed to identify harms that could result from use of the system, including through reasonably foreseeable misuse. Measures must be taken to mitigate the risk of harm.	AIDA may introduce new regulations regarding data usage in the context of AI. Expect new security requirements in AI tooling, and make sure that current and prospective vendors answer questions about the security of their AI usage, including basic controls such as data security, asset management, and identity and access management.
Accountability	Accountability means that organizations must put in place governance mechanisms needed to ensure compliance with all legal obligations of high-impact AI systems in the context in which they will be used. This includes the proactive documentation of policies, processes, and measures implemented.	New regulations are likely, prompting companies to inquire with third parties about compliance with any emerging reporting requirements and mandates.
Validity & Robustness	Validity means a high-impact AI system performs consistently with its intended objectives. Robustness means a high-impact AI system is stable and resilient in a variety of circumstances.	Organizations should ask their third parties about any validity issues with AI models in their operations, a concern that may be relevant to technology vendors and potentiallyextend to the physical supply chain.

Ultimately, the Canadian government is taking a hard look at how to regulate AI usage nationwide. There are going to be new mandates and new laws to comply with, no matter what. So, it makes sense for companies doing business in Canada or working with Canadian companies to understand any upcoming requirements as AIDA comes closer to passage.

Global Framework: ISO 42001

Artificial Intelligence (AI) is transforming how organizations operate, creating new efficiencies while introducing unique governance, risk, and compliance (GRC) challenges. Because of this, organizations must understand how to develop the necessary AI governance policies to use the technology safely and securely. ISO/IEC 42001:2023 is the first global standard for establishing an Artificial Intelligence Management System (AIMS), designed to help organizations address concerns around AI ethics, transparency, bias, safety, and privacy.

What is ISO 42001?

ISO/IEC 42001:2023 is an international standard developed by the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC) that provides guidance for establishing, implementing, maintaining, and continually improving an AIMS. It includes controls for ethical use, data quality, transparency, accountability, and third-party oversight. Like all ISO standards, ISO 42001 is voluntary, but it is quickly becoming a global benchmark for AI governance.

ISO 42001 Overview

ISO/IEC 42001 is modeled on the Plan-Do-Check-Act (PDCA) cycle and shares a common structure with other ISO management system standards like ISO 27001. ISO 42001 is structured with detailed annexes and core clauses that provide guidance for implementation. The structure encourages integration with other compliance frameworks and supports cross-functional governance.

Scope & Applicability

ISO 42001 applies to AI providers, producers, and users of AI systems regardless of origin or use case. AI used via SaaS, third-party APIs, or internal models can all fall within scope. The standard also extends to non-technical departments using AI tools (e.g., marketing) as part of the evaluation.

Alignment with Other Standards/Regulations

ISO 42001 harmonizes with ISO 27001 (Information Security), ISO 27701 (Privacy), and ISO 23894 (AI Risk Management). It complements multiple global frameworks and regulations, including but not limited to the NIST AI RMF, the EU AI Act, and DORA, allowing a single governance structure to serve multiple frameworks.

Why ISO 42001 Matters for Risk Management Teams

AI introduces various types of risks — from bias in algorithms and data misuse to regulatory exposure and reputational harm. The widespread use of large language models (LLMs), such as Claude, ChatGPT, and Gemini, is already creating new vulnerabilities, unearthing risks that IT security teams scramble to navigate. ISO 42001 provides a future-ready, auditable framework to: AI ethics, transparency, bias, safety, and privacy.

- Manage AI risk across the AI system's lifecycle
- Embed ethical principles and transparency
- Comply with regulations like the EU AI Act
- Build trust with stakeholders

ISO 42001 & Third-Party Risk Management

ISO/IEC 42001 fundamentally expands the scope of third-party risk management programs by formalizing AI-specific requirements across the supplier lifecycle. It requires organizations to assess and manage not just traditional security and compliance risks, but also the ethical, operational, and societal risks introduced by vendors' AI systems.

For TPRM teams, this means extending due diligence questionnaires to evaluate a supplier's AI governance maturity, monitoring ongoing changes to third-party AI models, and embedding contractual provisions that enforce transparency, explainability, and incident response obligations.

Going forward, TPRM programs must treat AI vendors as critical control points, integrating ISO 42001-aligned criteria into onboarding, monitoring, and audit processes to meet growing regulatory demands and ensure AI supply chain resilience.

Mapping Mitratech TPRM Capabilities to ISO 42001 Standards

The summary table below maps capabilities in the Mitratech Third-Party Risk Management Platform to select third-party, vendor, and supplier controls present in ISO 42001.

Note: *This table should not be considered definitive guidance. For a complete list of controls, please review the complete ISO standards in detail and consult your auditor.*

ISO 42001 Controls	How Mitratech Helps
Clause 6.1: Actions to address risks and opportunities	
6.1.2 AI risk assessment *"Organizations shall determine the risks and opportunities that need to be addressed to give assurance that AI management systems can achieve their results, prevent or reduce undesired effects, and achieve continual improvement."* *"The organization shall define and establish an AI risk assessment process that analyses the AI risks to:* *1) assess the potential consequences to the organization, individuals and societies that would result if the identified risks were to materialize;* *2) assess, where applicable, the realistic likelihood of the identified risks;* *3) determine the levels of risk"*	Mitratech partners with you to build a comprehensive third-party risk management (TPRM) program aligned with your broader artificial intelligence management systems program, based on proven best practices and extensive real-world experience. As part of the process, Mitratech can help to define: • Risk scoring and thresholds based on your organization's risk tolerance • Assessment and monitoring methodologies based on third-party criticality • Risk mitigation and remediation strategies
Annex A Control A.10 – Third-party and Customer Relationships	
A.10.2 Allocating responsibilities *"The organization shall ensure that responsibilities within their AI system lifecycle are allocated between the organization, its partners, suppliers, customers and third parties."*	Our experts collaborate with your team on defining and implementing TPRM processes and solutions; selecting risk assessment questionnaires and frameworks; and optimizing your program to address the entire third-party risk lifecycle – from sourcing and due diligence to termination and offboarding – according to your organization's risk appetite. As part of this process, Mitratech can help you define: • Clear roles and responsibilities (e.g., RACI) • Third-party inventories • Risk scoring and thresholds based on your organization's risk tolerance

	• Assessment and monitoring methodologies based on third-party criticality • Fourth-party mapping • Sources of continuous monitoring data (cyber, business, reputational, financial) • Key performance indicators (KPIs) and key risk indicators (KRIs) • Governing policies, standards, systems, and processes to protect data • Compliance and contractual reporting requirements against service levels • Incident response requirements • Risk and internal stakeholder reporting • Risk mitigation and remediation strategies
A.10.3 Suppliers *"The organization should establish a process to ensure that its usage of services, products or materials provided by suppliers aligns with the organization's approach to the responsible development and use of AI systems."*	Mitratech standardizes assessments against ISO best practices and other information security control frameworks, providing internal audit and IT security teams with a central platform for measuring and demonstrating adherence to secure software development and software development lifecycle (SDLC) practices. For organizations with limited resources and expertise, Mitratech can manage the third-party risk lifecycle on your behalf – from onboarding suppliers and collecting evidence, to providing remediation guidance and reporting on contract SLAs. As a result, you reduce vendor risk and simplify compliance without burdening internal staff.

Best Practice Recommendations for TPRM Leaders

The following list offers practical strategies for managing risks, ensuring transparency, and aligning programs with ISO 42001 standards.

- ☐ Clearly define the scope: internal AI use (e.g., usingChatGPT) vs. AI in customer-facing products (e.g., in-house models)
- ☐ Use Statements of Applicability (SOAs) to transparently document what's in-scope and how it's controlled
- ☐ Implement consistent risk assessment criteria for evaluating third-party AI systems
- ☐ Verify that third-party AI algorithms are accountable, fair, and free from bias
- ☐ Monitor and improve third-party AI systems through regular evaluations and updates
- ☐ Assess and manage risks associated with data used by third-party AI systems, focusing on data quality, privacy, and security
- ☐ Ensure that third-party AI systems adhere to ethical principles and maintain transparency in their operations
- ☐ Add AI scoping questions to your intake form to capture model type, data sensitivity, and decision criticality
- ☐ Publish an AI supplier code of conduct aligned with ISO 42001 Annex A principles (fairness, transparency, privacy, security)
- ☐ Require a Statement of Applicability or ISO 42001 certificate from critical AI vendors by a defined deadline
- ☐ Integrate AI incident clauses into your existing breach notification SLA
- ☐ Review the supplier tiering model quarterly to capture changes in AI usage

By embedding these enhancements, your TPRM program will not only satisfy ISO 42001 auditors but also provide a forward-looking defense against the unique, rapidly evolving risks that AI introduces into the extended enterprise. Look for solutions that offer automated evidence collection, continuous monitoring capabilities, and map risk assessment to relevant frameworks and regulations to satisfy auditors, reduce risk exposure, and stay compliant.

Looking Ahead: Turning These Insights into Action

ISO 42001 is a foundational AI governance standard increasingly seen as essential, transitioning from a "nice-to-have" to a baseline requirement for any AI-driven product strategy.

It requires structured risk and compliance processes that span the entire AI lifecycle, encompassing ethics, third-party oversight, and performance management. Risk and impact assessments must expand beyond classic cybersecurity and move into model fairness, data provenance, and human oversight as core considerations. Continuous vendor monitoring is mandatory; hidden AI in third-party tools (think shadow AI and fourth-party AI dependencies) is now a top governance gap.

Risk leaders should act now to future-proof their organizations' AI strategies and demonstrate trust, accountability, and compliance. Certification is less about a compliance badge and more about accelerated market trust amid tightening regulation.

Incorporating ISO 42001 into third-party risk management programs enhances an organization's ability to govern AI responsibly, mitigate potential risks, and foster confidence among stakeholders.

Discover how Mitratech's risk management platform can streamline AI governance and manage vendor AI risks. Request a demo of our third-party risk management solutions today.

Book a Demo ▶

Best Practices for Third-Party Risk Management Programs

Building questions about artificial intelligence into third-party risk management programs is imperative as the technology becomes more prominent now and in the future. The most effective way to do this is to find relevant guidance from standards bodies or regulators and use that information as building blocks to form a more cohesive program.

One of the most comprehensive forms of guidance comes in the form of the NIST AI Risk Management Framework. The AI RMF provides four key categories for risk managers seeking to place AI risk in context. These are Govern, Map, Measure, and Manage. Each category has distinct controls that are relevant to different components of an integrated risk management process.

The NIST AI RMF then breaks down its four core functions into 19 categories and 72 subcategories that define specific actions and outcomes. (NIST also offers a handy playbook that further explains the actions.)

Key TPRM-specific categories in the RMF include:

- **GOVERN 6** - Establish governing policies, standards, systems, and processes to protect data and systems from AI risks as part of your overall TPRM program
- **MAP 4** - Profile and tier third parties and quantify the inherent risks that third-party AI usage introduces to your organization to ensure that all risks are mapped
- **MEASURE 1-4** - Conduct comprehensive third-party risk assessments and continuously monitor and measure AI-specific risks in the context of your TPRM program
- **MANAGE 3** - Ensure comprehensive incident response to AI-specific risks from third-party entities

The table on the next page includes the four functions and select categories in the framework and suggests considerations to address potential third-party AI risks.

Note: *This is a summary table. For a full examination of the NIST AI Risk Management Framework, download the full version and engage your organization's internal audit, legal, IT, security, and vendor management teams.*

NIST AI RMF Category	TPRM Considerations
Govern is the foundational function in the RMF that establishes a culture of risk management; defines processes; and provides structure to the program.	
GOVERN 1: Policies, processes, procedures, and practices across the organization related to the mapping, measuring, and managing of AI risks are in place, transparent, and implemented effectively.	Build AI policies and procedures as part of your comprehensive third-party risk management (TPRM) program in line with your broader information security and governance, risk, and compliance frameworks. Seek out experts to collaborate with your team on defining and implementing AI and TPRM processes and solutions; selecting risk assessment questionnaires and frameworks; and optimizing your program to address AI risks throughout the entire third-party lifecycle – from sourcing and due diligence, to termination and offboarding – according to your organization's risk appetite. As part of this process, you should define: • Governing policies, standards, systems, and processes to protect data from AI risks. • Legal and regulatory requirements, ensuring that third parties are assessed accordingly. • Clear roles and responsibilities (e.g., RACI) for team accountability. • Risk scoring and thresholds based on your organization's risk tolerance. • Assessment and monitoring methodologies based on third-party criticality are continually reviewed. • Third-party AI inventories. • Fourth-party mapping to understand exposure to AI usage-based risks in your extended ecosystem. • Key performance indicators (KPIs) and key risk indicators (KRIs) for internal stakeholders. • Contractual requirements and right to audit. • Incident response requirements. • Risk and internal stakeholder reporting. • Risk mitigation and remediation strategies.
GOVERN 2: Accountability structures are in place so that the appropriate teams and individuals are empowered, responsible, and trained for mapping, measuring, and managing AI risks.	
GOVERN 3: Workforce diversity, equity, inclusion, and accessibility processes are prioritized in the mapping, measuring, and managing of AI risks throughout the lifecycle.	
GOVERN 4: Organizational teams are committed to a culture that considers and communicates AI risk.	
GOVERN 5: Processes are in place for robust engagement with relevant AI actors.	
GOVERN 6: Policies and procedures are in place to address AI risks and benefits arising from third-party software and data and other supply chain issues.	

NIST AI RMF Category	TPRM Considerations
Map is the function that establishes the context to frame risks related to an AI system.	
MAP 1: Context is established and understood.	Developing a good risk management process and understanding the context of AI usage begins with profiling and tiering third parties, which involves quantifying the inherent risks for all third parties, specifically the inherent AI risks. Criteria used to calculate inherent risk for third-party classification and categorization includes: • Type of content required to validate controls. • Criticality to business performance and operations. • Location(s) and related legal or regulatory considerations. • Level of reliance on fourth parties (to avoid concentration risk). • Exposure to operational or client-facing processes. • Interaction with protected data. From this inherent risk assessment, your team can automatically tier suppliers according to AI risk exposure; set appropriate levels of further diligence; and determine the scope of ongoing assessments. Rule-based tiering logic enables vendor categorization using a range of data interaction and regulatory considerations.
MAP 2: Categorization of the AI system is performed.	
MAP 3: AI capabilities, targeted usage, goals, and expected benefits and costs compared with appropriate benchmarks are understood.	
MAP 4: Risks and benefits are mapped for all components of the AI system including third-party software and data.	
MAP 5: Impacts to individuals, groups, communities, organizations, and society are characterized.	

NIST AI RMF Category	TPRM Considerations
Measure is the function that analyzes, assesses, benchmarks, and monitors AI risk and related impacts.	
MEASURE 1: Appropriate methods and metrics are identified and applied.	Look for solutions that feature a large library of pre-built templates for third-party risk assessments. Third-party vendors should be evaluated for their AI practices at the time of onboarding, contract renewal, or at any required frequency (e.g., quarterly or annually), depending on material changes. Assessments should be managed centrally and backed by workflow, task management, and automated evidence review capabilities to ensure that your team has visibility into third-party risks throughout the relationship lifecycle. Importantly, a TPRM solution should include built-in remediation recommendations based on risk assessment results to ensure that your third parties address risks in a timely and satisfactory manner and can provide the appropriate evidence to auditors. To complement vendor AI evaluations, continuously track and analyze external threats to third parties. As part of this, monitor the Internet and the dark web for cyber threats and vulnerabilities. All monitoring data should be correlated with assessment results and centralized in a unified risk register for each vendor, streamlining risk review, reporting, remediation, and response initiatives. Monitoring sources typically include: • 1,500+ criminal forums; thousands of onion pages; 80+ dark web special access forums; 65+ threat feeds; and 50+ paste sites for leaked credentials — as well as several security communities, code repositories, and vulnerability databases covering 550,000 companies. • Databases containing 10+ years of data breach history for thousands of companies around the world. Finally, continuously measure third-party KPIs and KRIs against your requirements to help your team identify risk trends, determine the status of third-party risks, and pinpoint exceptions to common behavior that may warrant further investigation.
MEASURE 2: AI systems are evaluated for trustworthy characteristics.	
MAP 3: Mechanisms for tracking identified AI risks over time are in place.	
MEASURE 4: Feedback about efficacy of measurement is gathered and assessed.	

NIST AI RMF Category	TPRM Considerations
The **Manage** function entails allocating risk resources to mapped and measured risks on a regular basis and as defined by the GOVERN function. This includes plans to respond to, recover from, and communicate about incidents or events.	
MANAGE 1: AI risks based on assessments and other analytical output from the MAP and MEASURE functions are prioritized, responded to, and managed.	As part of your broader incident management strategy ensure that your third-party incident response program enables your team to rapidly identify, respond to, report on, and mitigate the impact of third-party vendor AI security incidents. Key capabilities in a third-party incident response service include: • Continuously updated and customizable event and incident management assessments. • Real-time questionnaire completion progress tracking. • Defined risk owners with automated chasing reminders to keep surveys on schedule. • Proactive vendor reporting. • Consolidated views of risk ratings, counts, scores and flagged responses for each vendor. • Workflow rules to trigger automated playbooks to act on risks according to their potential impact on the business. • Built-in reporting templates for internal and external stakeholders. • Guidance from built-in remediation recommendations to reduce risk. • Data and relationship mapping to identify relationships between your organization and third, fourth or Nth parties to visualize information paths and reveal at-risk data. • Armed with these insights, your team can better manage and triage third-party entities, understand the scope and impact of the incident, determine which data was involved, whether the third party's operations were impacted, and when remediation has been completed.
MANAGE 2: Strategies to maximize AI benefits and minimize negative impacts are planned, prepared, implemented, documented, and informed by input from relevant AI actors.	
MANAGE 3: AI risks and benefits from third-party entities are managed.	
MANAGE 4: Risk treatments, including response and recovery, and communication plans for the identified and measured AI risks are documented and monitored regularly.	

Mapping third-party risk considerations onto the NIST AI RMF is only the first step in the process, of course. Following this initial work, TPRM program managers should follow their normal due diligence workflow to profile and tier vendors based on their AI capabilities and decide on appropriate remediations based on their own risk tolerances.

The Future of AI Regulations and Third-Party Risk Management

Governments around the world are actively debating how to regulate artificial intelligence technology. Common concerns include responsible development to limit societal impacts, data privacy concerns, and potential software supply chain exposures. Regulators have also looked at limiting specific use-cases, suggesting that the regulations arising around the world will likely combine privacy, security, and ESG concerns.

The next few years will likely offer more clarity on how organizations worldwide need to adapt their third-party risk management programs to AI technology. So far, the EU remains the only jurisdiction with a comprehensive AI regulation that has an enforcement mechanism. Whether that will remain true or not in the next few years is anyone's guess. What is accurate, however, is that companies are rapidly integrating AI into their operations and governments will eventually need to provide legal frameworks. At this point, adopting a more cautious and considerate approach to AI in operations and asking vendors and suppliers questions about their approach and risk management is the correct choice for third-party risk managers.

How Mitratech Can Help

Mitratech can help your organization improve not only its own AI governance, but also how it governs third-party AI risks. Specifically, we can help you:

- Establish governing policies, standards, systems and processes to protect data and systems from AI risks as part of your overall TPRM program. (Aligns with AI RMF category GOVERN 6.)
- Profile and tier third parties, while quantifying inherent risks associated with third-party AI usage to ensure that all risks are mapped. (Aligns with AI RMF category MAP 4.)
- Conduct comprehensive third-party risk assessments and continuously monitor and measure AI-specific risks in the context of your TPRM program. (Aligns with the AI RMF MEASURE category.)
- Ensure comprehensive incident response to AI-specific risks from third-party entities. (Aligns with AI RMF MANAGE 3.)

Leveraging the NIST AI Risk Management Framework in your TPRM program will help your organization establish controls and accountability over third-party AI usage. As the regulatory environment changes over time, being able to account for AI changes and leverage the technology effectively will be increasingly vital.

For more on how Mitratech can help simplify this process, request a demo today.

Book a Demo ▶

MITRATECH

About Mitratech

Mitratech is a proven global technology partner for legal, risk, compliance, and HR professionals seeking to maximize productivity, control expense, and mitigate risk by deepening operational alignment, increasing visibility, and spurring collaboration across an enterprise. Mitratech serves over 24,000 organizations worldwide, spanning more than 160 countries.

Learn more at Mitratech.com.

EMPOWER. AUTOMATE. ELEVATE.