onetrust | protiviti® Global Business Consulting

Establishing a scalable AI governance framework:

Key steps and tech for success

EBOOK | 2024

onetrust

protiviti®

Global Business Consulting

Table of Contents

Strategic AI governance 3

Part 1: Establishing a structure 4

- How does an organization start to build out its AI inventory? 4
- What are some of the considerations when establishing responsible AI principles and what are the core components of an AI policy? 7
- What are some tests and risk assessments organizations should be doing? 9
- How does an organization bring their AI policies to life? 11
- What guidance do laws and frameworks offer? 12
- Common controls and best practices for governing AI 13

Part 2: Understanding the AI 14

- What are the required technical capabilities of an AI governance program? 14
- Model risk management 15
- Data privacy management 16

About 17

Strategic AI governance

Balancing risk, compliance, and innovation

When it comes to establishing an operating model for AI governance, which includes policies, development standards, and risk management approaches, organizations often face a classic chicken or egg dilemma. Should they first identify and understand the AI technologies and use cases they have before establishing governance structures, or should they establish governance mechanisms first and establish AI use cases in compliance with that structure?

The answer is subjective to the organization, its maturity, and culture; regardless, taking the steps outlined in this book will help you recognize the required actions necessary for your organization to identify, manage, and measure AI solutions and their effectiveness of AI solutions to ensure continuous performance and appropriate risk management are in place as part of an overall governance approach.

Part 1: Establishing a structure

If your organization decides to first establish your governance structure, there are a few key questions you can ask yourself to begin this process. Even if your organization doesn't choose to establish a governance structure as the first step, these considerations can be applied regardless of where you are in the AI lifecycle when the need to formalize a governance structure arises.

How does an organization start to build out its AI inventory?

Creating an inventory of AI systems and use cases is crucial for developing an effective AI governance strategy and managing risks. This process allows organizations to gain a clear understanding of their existing AI technologies, ongoing AI initiatives, and future plans.

Once they design and operationalize an intake process for new uses of AI and identify the AI technologies they already have and are accountable for, the inventory can help them understand the scale, complexity, desired outcomes, and potential risks associated with their AI initiatives.

Here are some steps the organization can follow to begin to build out their AI inventory:

Define terminology and taxonomy for AI components
Given the current lack of consensus regarding the definition of AI, it's important for an organization to establish their own clear definition of AI-related terms. Having these definitions to apply to their unique business and risk environment will help the organization support consistent identification of AI technologies – including third parties using AI on their behalf – and enable informed risk management decisions. Clear terminology supports systematic and consistent identification of AI throughout the organization, as well as the use of AI by its partners as part of its third-party risk management function.

When developing a strategy and approach to governance, the organization must consider their mission and goals, organizational objectives and key results, industry sector, risk tolerance, regulatory environment, and the current technology landscape. Understanding and defining the scope of AI use and identifying the attributes of AI that will factor into risk management decisions – such as functional application of AI, data consumed, open vs. closed-source AI, third-party models, and type of AI (e.g. GenAI, ML, NLP, etc.) – is the critical first step for an organization in standing up their AI inventory.

Identify all current and proposed AI systems and use cases
Once the terminology is defined, you can begin mapping out all the instances of AI currently in use, being evaluated, or those planned for implementation across the organization. This mapping exercise can include specific platforms or applications with AI components, GenAI models, and algorithms.

This will involve reaching out to other stakeholders within the organization, including but not limited to IT, operations, sales, marketing, and HR, to ensure all business functions using or impacted by AI are identified. Automated scanning tools and integrations with your ML tech stack can assist in this effort by identifying ML models within developer environments and uncovering AI applications across your organization's ecosystem.

Part 1: Establishing a structure

Part of identifying the AI in use includes reducing the risk of the use of "shadow AI" by employees and partners. Specifically, shadow AI refers to AI that is unknown, untracked, and unmanaged by IT and risk management functions.

End-users can be resourceful when it comes to finding and accessing AI solutions that could result in intellectual property leaks, misuse of both the AI and its output, and legal and regulatory exposure.

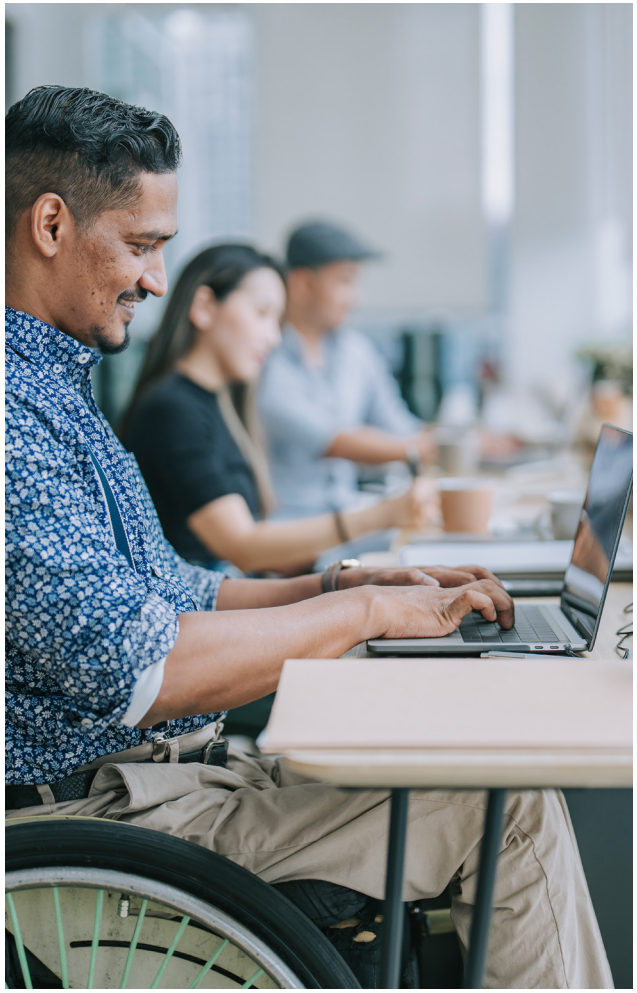
Document key information for each use case
For each discovered application of AI, collect essential details as defined by your developed taxonomy, including but not limited to the purpose, capabilities, methods of processing (use of generative AI, automated decision-making, etc.), inputs (the data used), solution features, outputs (decisions made or tasks performed) and training methods of your system.

You should also collect information about the developers (whether an internal team or external vendor), users (internal staff, customers, partners, or other external stakeholders), deployment environment (on-premises/cloud/hybrid), and regulatory compliance requirements in the AI inventory.

In addition, for each AI use, there should be at least a high-level description of the problem it solves or what value it brings to the organization, and for larger investments, there should be a fully documented business case.

Classify AI use cases based on risk rating
Once AI use cases are documented along with their business case, it's possible and advisable to classify them based on their risk rating. It is important that risk criteria are established early on to identify high-risk uses in particular, or proposed use cases that could expose the organization to risks beyond its tolerance.

After risk criteria are established – typically with the direct involvement of legal, risk, and compliance functions – the risk rating can be determined by factors such as business value, inherent operational risk, residual risk after the application of controls, likelihood of occurrence, potential impact on business operations if things go wrong, sensitivity of data used, legal implications, or reputational risk. Once use cases are risk-ranked, they can be risk-prioritized for review and remediation.



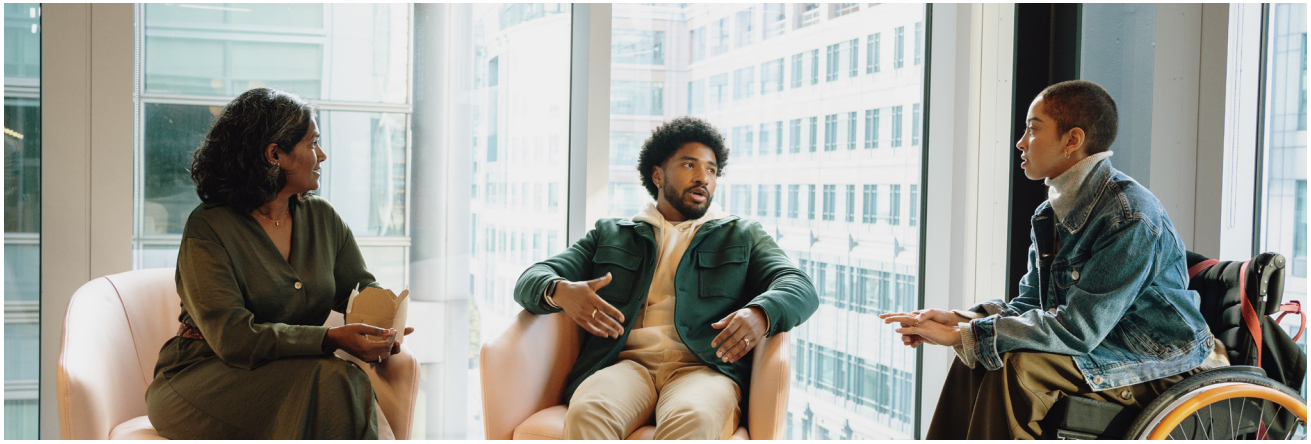
Part 1: Establishing a structure

In addition to maintaining awareness of AI-specific regulatory risks that apply to them, organizations should ensure their AI risk tolerance and risk ranking system are well integrated with the enterprise risk management function overall. This will help prioritize governance efforts towards high-risk use cases first.

Maintain a living inventory

The AI inventory is not a one-time effort, but rather should be updated continuously as new use cases are introduced, as existing ones evolve or retire, or as changes are required within the AI solutions that are already deployed.

This approach ensures that the inventory remains current and relevant ("evergreen") and can be relied on to support ongoing governance and continuous risk management efforts. With each material change you can invoke testing, evaluation, validation, and verification (TEVV), you learn more about the impacts and risks of the specific AI use case, and this new information can be incorporated into the inventory.



Certain triggers for material changes, like a significant feature change determined by AI developers, or a significant regulatory update identified by risk and compliance functions, can be defined and used to initiate a new TEVV. For example, if an AI use case is planned to be used for sensitive information processing, a data protection impact assessment (DPIA) can and should be automatically triggered.

Integrate with existing technology inventories

Integrate the AI inventory with existing technology inventories within the organization, such as the central IT inventory of systems and components, or the current record of processing activities used to track business processes leveraging personal information, for a holistic view of how AI fits into the larger tech ecosystem.

Creating an inventory frequently requires considerable cross-functional collaboration, so having leadership buy-in is key to success in this venture.

Part 1: Establishing a structure

What are some of the considerations when establishing responsible AI principles and what are the core components of an AI policy?

Several key considerations should be integrated into the AI policy to ensure that the deployment and use of AI technologies adheres to responsible and ethical principles, legal and regulatory requirements, and leading practices for development, deployment, and use, including embedding risk mitigation as part of design.

The core components of an AI policy for an organization can be distilled into the following elements that will help create a strong foundation for responsible AI use that not only complies with existing regulations but also promotes trust among stakeholders by demonstrating a commitment to ethical practices:

AI principles

Identify the governance framework(s) that your organization already uses (e.g., privacy, security, data governance), and the AI governance frameworks you plan to use as reference for responsible use and management of AI (EU AI Act, NIST AI RMF, Microsoft Responsible AI Principles, etc.).

Consider the regulatory landscape as well as industry standards for supplemental guidance around AI principles, and use these to define and document principles of AI aligned to the broader goals and initiatives of your organization. These might include principles like fairness, transparency and explainability, accountability, security, and respect for human rights, among others.

Policy statement

This is a declaration of the organization's commitment to responsible AI use. It should outline the policy's purpose and highlight the approach to the ethical, legal, and societal aspects of AI.

Leadership and governance structure

Define how the established AI principles will be operationalized and enforced throughout the organization, and detail the roles and responsibilities for managing these efforts within it, including how the lines of defense will work in concert to identify and mitigate risk.

This includes oversight committees or boards, general and role-specific education and awareness programs (e.g., all-employee, IT-specific, third-party training and reference materials) as well as their composition and functions in relation to AI governance. Organizations may update existing enterprise policies to include AI considerations; they may also choose to centralize AI into a single AI policy.

Part 1: Establishing a structure

Compliance and risk management strategy
Develop the compliance and risk management strategies for meeting with current regulatory requirements, identifying and remediating AI risk, safeguarding input/output data, and implementing performance and compliance monitoring mechanisms. A compliance and risk management strategy can also be used to develop AI training for the organization, manage stakeholder engagement, and maintain continuous improvement.

Include robust practices around data privacy, security, quality control, and consent mechanisms to ensure proper management throughout the data lifecycle from collection to disposal. Specify procedures to ensure adherence with all relevant laws, regulations, and industry standards concerning AI use. This could involve regular compliance checks or audits.

Explain how potential risks associated with AI development and deployment are identified and mitigated; discuss strategies for managing these risks, including technical safeguards and policy enforcement measures.

These components build upon one another to create a comprehensive policy that not only guides internal operations but also communicates the organization's commitment to responsible AI usage to external stakeholders, such as clients or regulatory bodies.

Part 1: Establishing a structure

What are some tests and risk assessments organizations should be doing?

To understand their AI risk posture and level of AI risk, organizations should be conducting a variety of tests and risk assessments. Here are some key ones:

Data quality checks
Implement robust data governance practices that address data quality checks. Incomplete, stale, bad or biased data can lead to faulty decisions by an AI system, leading to inaccurate, incomplete, or incorrect responses.

AI risk assessment
Perform comprehensive risk assessments focusing on identifying potential harm from the use of AI and ensuring alignment with established AI principles. This involves identifying and evaluating risks associated with each use case, such as data privacy or regulatory compliance concerns, potential for algorithmic bias leading to unfair outcomes, security vulnerabilities, etc.

It's also important to assess risks associated with not just operational aspects but also strategic ones, like financial or reputational damage due to unethical AI usage or non-compliance with emerging regulations around AI.

Furthermore, it is important that risk is continuously reassessed, not only for specific use cases (following intake, following testing and validation, and following implementation to account for changes), but also for the organization as a whole in the form of a "risk appetite" statement; namely, the determination and description of how much risk the organization is willing to take on and which risks it will avoid by design.

This allows the organization to balance individual business line choices with its overall enterprise risk strategy, which includes protecting its brand and reputation, its stock price, and its regulatory compliance posture.

Privacy impact analysis and DPIA
A privacy impact assessment (PIA) helps organizations that handle personal information ensure they comply with data protection laws and regulations. It also helps organizations identify and mitigate potential risks associated with processing personal information.

Similarly, a DPIA can be conducted to analyze, identify, and minimize risks related to safeguarding all data processed by the organization, regardless of whether the data contains personal information.

AI red team testing
AI red team testing is an ethical hacking approach where a group mimics potential adversarial attacks to evaluate the robustness and security of an AI system. Testing is conducted within controlled environments by professionals with expertise in cybersecurity and AI. This practice helps organizations uncover vulnerabilities in the AI, allowing the organization to understand real threats and improve existing mitigation processes.

Part 1: Establishing a structure

Performance and compliance monitoring
Establish ongoing performance monitoring processes to regularly evaluate how uses of AI functionality are working in real-world scenarios post-deployment compared to during development/testing phases. Another important consideration for monitoring AI performance continuously is inclusion of a feedback mechanism for those impacted by AI to report issues with inaccurate or harmful outputs.

Similarly, compliance monitoring mechanisms should be established to track changes in the regulatory environment and trigger updates to program documentation and processes when necessary. Controls for continuous performance monitoring and compliance for AI use cases should be established following the risk assessments conducted at intake.

Crisis simulation exercises
Conduct crisis simulation exercises (tabletop exercises, playbook walkthroughs, etc.) where hypothetical situations involving failures, compromises, or breaches related to AI are acted out in order to test organization's preparedness for managing such incidents effectively. Organizations may integrate these exercises with their existing business continuity and disaster recovery program.

Part 1: Establishing a structure

How does an organization bring their AI policies to life?

Operationalizing an organization's AI policies involves several steps that ensure the policy is not just on paper, but also reflected in the actions and behaviors of the organization. Here's how this can be achieved:

Leadership commitment
First and foremost, for any policy to take effect, leadership at the C-suite level must commit to it. This means leaders need to visibly support the policy, set expectations for its implementation, and hold themselves and others accountable for following it.

Governance bodies and decision architecture
Organizations often have an executive and/or risk committee that will evaluate major investments prior to greenlighting them. For more minor investments that IT can make on behalf of the business, there may be smaller IT committees, like an IT risk committee. But there may also be some kind of investment governance body that includes both business interests and representation from risk management functions.

Outside of IT, there are other control-function established governance structures you can consider (including privacy, security, and data). Many organizations have decided to establish AI councils or working groups as the decision-making team for AI. All these need to be considered when forming your governance body, and regardless of which teams you decide to include, there needs to be a decision flow design that supports rather than inhibits accelerated responsible AI adoption.

Roles and responsibilities
Once in-scope governing bodies are identified, a clear RACI matrix and decision flow should be defined and documented. This should outline clear sponsorship, ownership, stewardship, and oversight of AI solutions (including at third parties/solutions used by them).

Often, these bodies are divided into two camps: those reviewing the business case at a high level and those evaluating the functionality and risk mitigation approach at a more technical level. Regardless, all these boxes must be checked for AI solutions to be managed by appropriate personnel on an ongoing basis.

Policy communication
The AI policy should be clearly communicated across the organization through various channels such as internal newsletters, meetings, training sessions, and the like. This ensures that everyone understands what is expected of them when using or developing AI systems. A routine leadership review of the policy should also be established.

Training programs
Develop comprehensive training programs to educate employees about the responsible use of AI technologies as outlined in organizational policies. These trainings should include practical examples of how to apply ethical principles in day-to-day work with AI systems. The training program should also include triggers for retraining on a routine basis, such as the time since the last training was completed or the score falling below an acceptable threshold.

Part 1: Establishing a structure

Feedback channels
Create channels for employees and other stakeholders to provide feedback on the policy implementation or raise concerns about potential ethical issues related to AI use. Implement processes and designate accountable parties for responding to feedback.

Iterative improvement
Review and update AI policies regularly based on feedback received, changes in technology or societal expectations, new regulations, etc., ensuring they remain relevant and effective over time. New and more powerful features are added to technology month to month, and organizations may need to consider whether an annual review is adequate or if a more frequent review is necessary.

External communication & reporting
Publicly communicate commitment to responsible use of AI through things like annual reports or website updates. This builds trust among external stakeholders –among them, clients, customers and or regulatory bodies – showcasing transparency in operations involving AI technologies.

Organizations can also regularly engage with customers, partners, regulators, etc., sharing insights about their approach to responsible use of AI technologies while also gaining perspectives from these stakeholders. This, in turn, can aid improvement in the organization's AI practices.

What guidance do laws and frameworks offer?

Current domestic and global AI laws and frameworks encompass several key areas, including:

Transparency and explainability
Organizations should strive to make their AI systems as transparent as possible. This involves explaining how these systems make decisions in a manner that users can understand. AI systems should not operate as "black boxes;" users have the right to know how these systems work, especially when decisions significantly impact them.

Data privacy
Protecting personal data used by AI systems is critical. Organizations must monitor regulatory changes so they can comply with data protection laws such as the GDPR in Europe or specific sector regulations elsewhere. This includes providing adequate data protection (both security and privacy) measures to prevent data breaches, obtaining necessary consents for data usage, and respecting users' rights over their personal data.

Risk management
Conduct thorough risk assessments regularly to identify potential risks associated with deploying AI technologies, including ethical risks like bias or discrimination, technical risks like system failures or cyber-attacks, and legal risks like non-compliance with regulations, and then develop mitigation strategies.

Continuous monitoring & improvement
Monitor the performance and compliance of deployed AI systems on an ongoing basis. This includes observing for any unintended consequences

Part 1: Establishing a structure

even after thorough testing has been conducted during the development phase and validating that all controls and guardrails are functioning as intended. Issues, incidents, and errors should be tracked and reported to relevant stakeholders, which may include notifying a supervisory authority if operating within the EU.

Stakeholder engagement and awareness
Engage stakeholders throughout the organization about the importance of responsible use of AI. Provide knowledge-share sessions focusing on understanding how the AI is deployed, what are its limitations, etc., which will help increase trust in these technologies among employees/customers alike while also aiding in better detection and reporting if something goes awry.

Common controls and best practices for governing AI based on current regulations and frameworks

Data quality controls
This includes processes and mechanisms for measuring accuracy, timeliness, completeness, and bias within input or output data.

Policy enforcement controls
Efforts should be made to establish enforcement mechanisms around the AI policy to ensure agreed-upon AI principles are followed; strive towards upholding principles/goals that global laws and regulations were designed around, even if there are no immediate penalties for non-compliance.

Security and privacy controls
Security and privacy controls are established to identify risks and vulnerabilities, as well as ensure the safeguarding of data and the ethical use of AI at every phase of the lifecycle. Specific controls may include monitoring for unauthorized access, misuse of information, and potential harm to individuals or organizations. Resilience exercises can also be integrated to measure the resilience of the AI system and determine whether offboarding from the model or use case is possible.

Monitoring controls
Establish controls for monitoring AI, including actual performance versus expected performance, human-in-the-loop involvement, and use of generative AI.

Remember, what matters most is not just having these controls in place but also ensuring they're effectively implemented across all layers of the organization – from strategic planning down to operational execution – which requires strong leadership commitment alongside active participation from all staff members regardless their role/function within company structure.

Part 2: Understanding the AI system

Once an effective governance structure is established to capture AI use cases more comprehensively, it will likely be necessary to operationalize the framework into an IT solution so that it can be scaled, the data captured is housed in a system that eases maintenance and use, and that key activities can be automated.

What are the required technical capabilities of an AI governance program?

The required technical capabilities of an AI governance program will allow you to dive deeper into the technical aspects of an AI system, ensuring that you have a complete understanding of the AI in use and how best to govern it. Here are some of the technical capabilities you want to have in place:

AI management system
AI management systems play a crucial role in the development and deployment of AI technologies. A robust AI management system will help with managing the planning, development, implementation, and deployment of AI systems – including risk evaluation, identification, management, and tracking across the various risk domains. Here are a few components your AI management system should have:

AI/ML risk identification and assessment
These assessments support identifying risks through automated means, rules-based logic, and human input. Once complete, you can use the results of these assessments to measure the impact of each AI risk and begin implementing controls to mitigate them.

AI/ML risk and control library
This library includes a list of risks and the proportionate mitigating controls informed by global AI laws and frameworks. It will help organizations respond more effectively if and when a risk or other breach occurs.

AI/ML risk and issue tracking
Have a documented workflow in place to assign tasks and collect evidence for identified risks.

AI/ML audibility and traceability
As mentioned before, transparency and explainability are key aspects of trustworthy AI and responsible use. Your AI management system should enable you to follow the flow of information through systems to understand the data inputs that helped a system arrive at an output.

System cards
Another aspect of transparency is ensuring plain language is used when describing an AI system and how it works. System cards provide explanations of an AI system's functions in easy-to-understand language so people interacting with it know its purpose and can choose to opt-out.

Part 2: Understanding the AI

Model risk management

Model risk management refers to managing the internal controls, audits, documentation, policies, and procedures for machine learning models. There are a few key aspects of this process to account for:

Model evaluations

These evaluations perform technical analysis of model performance against set standards for performance, accuracy, security, and fairness. Conducting these evaluations early and often in your AI lifecycle will help you catch any biases or inaccuracies in the AI model's outputs.

Model ops monitoring

This process includes monitoring machine learning operations (model ops) to ensure reliability of performance and compliance of the models. This entails monitoring ML models for changes such as model degradation, data drift, and concept drift, and ensuring the model maintains an acceptable level of performance. Types of model monitoring include reactive monitoring, proactive monitoring, real-time monitoring, log monitoring, performance monitoring, and security monitoring.

Model cards

Like system cards, model cards document key model details, including purpose, performance metrics, training data, and known limitations or biases.

Having an effective risk management framework is crucial for organizations that rely heavily on AI models for daily operations and decision-making; implementing these controls will help ensure your models don't provide you with biased, inaccurate, or otherwise harmful outputs.

AI application security

Although AI systems can be invaluable to organizations, they do pose unique security risks. These are some tools for developing, testing, and adding security features to applications to prevent vulnerabilities.

Adversarial resistance

Techniques and approaches for defending AI systems against adversarial attacks including adversarial training, defensive distillation, gradient masking, feature squeezing, randomized transformations, and ensemble techniques.

Security and TEVV teams can identify additional vulnerabilities not known by default so that controls (or mechanisms to address and mitigate the risks) are applied, maintained, and monitored.

For example, if you have a system that can sometimes produce erroneous output because of malicious tampering that is hard to detect by automation, having a human-in-the-loop control can assist with detecting the flawed output so that it can be addressed, tracked, and managed as a security breach incident.

Content anomaly detection

This process refers to finding patterns or instances in a dataset that deviate significantly from expected behavior. Catching these anomalies in your data early will help prevent undesired outcomes later.

Part 2: Understanding the AI

Data privacy management

At its core, managing AI requires managing data. To do this, continuous evaluation, monitoring, and auditing of AI systems to ensure compliance with data protection laws and ethical obligations is critical. Here are a few ways you can do this:

Data catalog

Create an inventory of data assets to facilitate the discovery, use, and protection of data for business use in an AI context.

Data lineage

The process of tracking the flow of data throughout its lifecycle, including its source, storage location, and any transformations, providing detailed visuals that highlight the relationship between downstream and upstream dependencies in the data pipeline.

Data policy enforcement

It's important to have mechanisms that ensure the confidentiality, integrity, and availability of data and documented policies to enforce those mechanisms.

PETs

Privacy-enhancing technologies (PETs) are tools that help protect personal data during storage, processing, and transmission to reduce the risks associated with data use. Some examples include homomorphic encryption, federated learning, pseudonymization, differential privacy, and synthetic data.

Establishing robust AI governance is not just about compliance; it's also about strategically managing risk and ensuring that AI initiatives drive meaningful value for the organization. By following the steps summarized in this book, steps—such as building a comprehensive AI inventory, establishing clear governance principles, and integrating risk assessments—organizations can create a governance framework that is both practical and scalable. This approach not only safeguards against potential pitfalls but also positions the organization to capitalize on AI's potential, ensuring that AI technologies are deployed responsibly, transparently, and in alignment with broader business objectives. As AI continues to transform industries, organizations that prioritize effective governance will be better equipped to navigate the challenges and opportunities ahead.

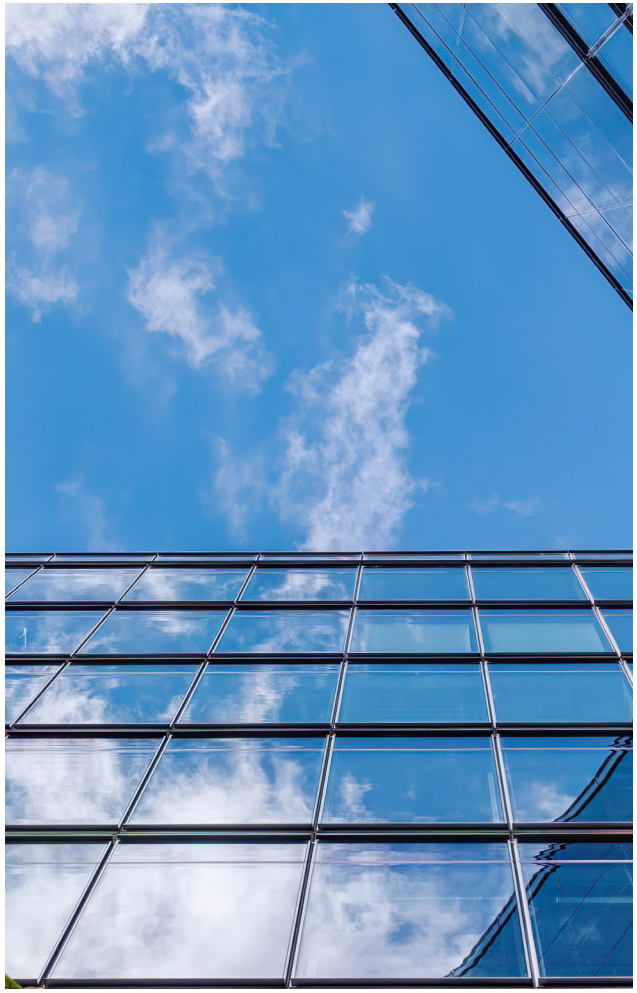
About

About OneTrust

OneTrust unlocks the full potential of data and AI, responsibly. Our platform enforces the secure handling of company data, empowering organizations to drive innovation responsibly while mitigating risks. With a comprehensive suite of solutions spanning data and AI security, privacy, governance, risk, ethics, and compliance, OneTrust enables seamless collaboration between data teams and risk teams to enable rapid and trusted innovation. Recognized as the market leader in trust, OneTrust boasts over 300 patents and serves more than 14,000 customers globally, ranging from industry giants to small businesses.

About Protiviti

Protiviti (www.protiviti.com) is a global consulting firm that delivers deep expertise, objective insights, a tailored approach and unparalleled collaboration to help leaders confidently face the future. Protiviti and its independent and locally owned member firms provide clients with consulting and managed solutions in finance, technology, operations, data, digital, legal, HR, risk and internal audit through a network of more than 90 offices in over 25 countries. Named to the Fortune 100 Best Companies to Work For® list for the 10th consecutive year, Protiviti has served more than 80 percent of Fortune 100 and nearly 80 percent of Fortune 500 companies. The firm also works with government agencies and smaller, growing companies, including those looking to go public. Protiviti is a wholly owned subsidiary of Robert Half Inc. (NYSE: RHI).



onetrust | protiviti Global Business Consulting

REQUEST A DEMO TODAY AT ONETRUST.COM

OneTrust unlocks the full potential of data and AI, responsibly. Our platform enforces the secure handling of company data, empowering organizations to drive innovation responsibly while mitigating risks. With a comprehensive suite of solutions spanning data and AI security, privacy, governance, risk, ethics, and compliance, OneTrust enables seamless collaboration between data teams and risk teams to enable rapid and trusted innovation. Recognized as the market leader in trust, OneTrust boasts over 300 patents and serves more than 14,000 customers globally, ranging from industry giants to small businesses.