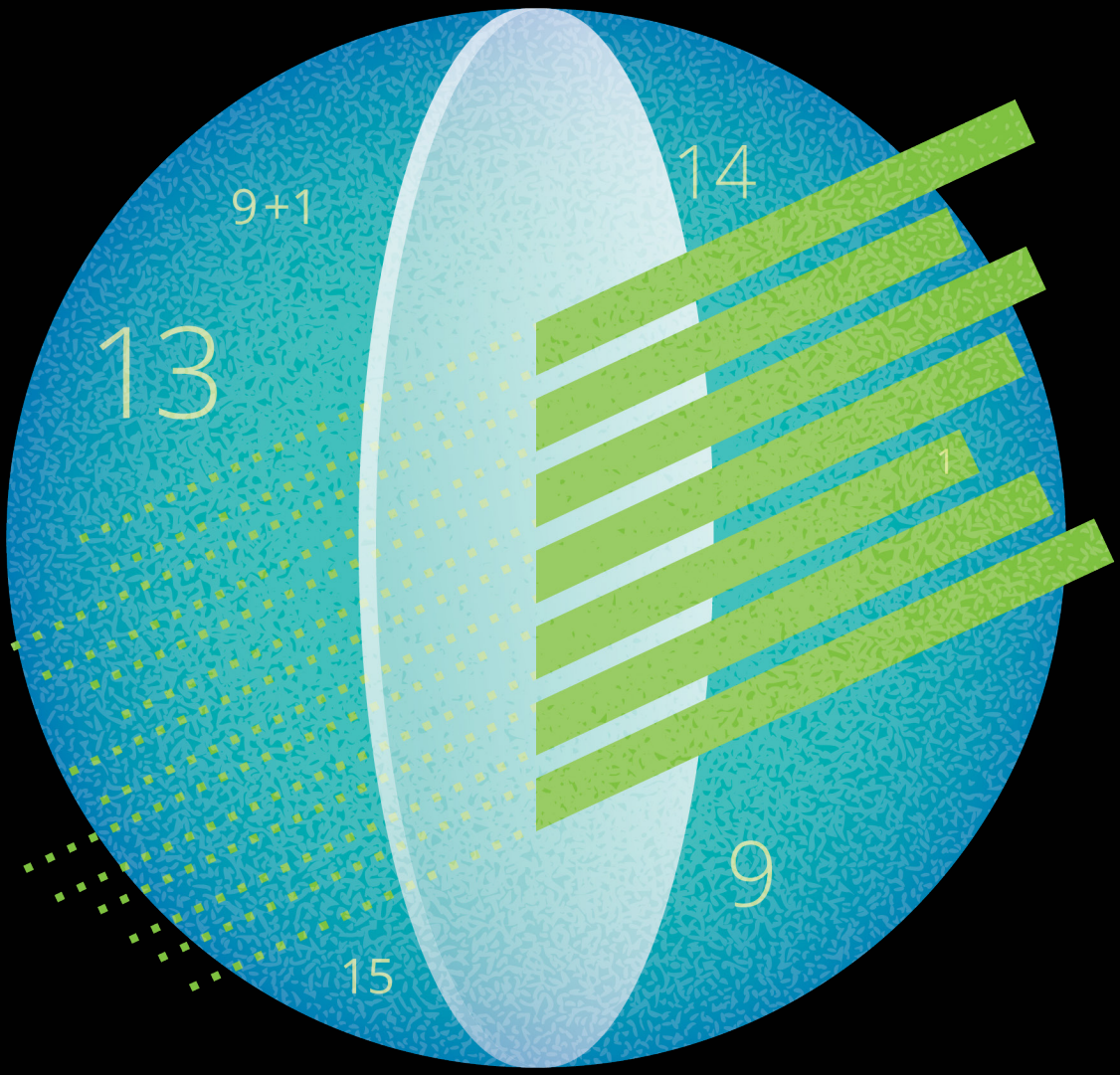
Deloitte.



OmniaAI

Building trust in AI

How to overcome risk and operationalize AI governance

Introduction

Artificial intelligence (AI) is viewed as a primary driver of innovation in almost every industry. Organizations have nevertheless been slow to embrace it because of the challenges and unknowns it introduces. Demystifying the risks inherent to AI is a key step in overcoming those challenges and more fully understanding how to extract AI's value.

Although the regulatory landscape is evolving, organizations can still begin to tackle AI risks. This requires robust, transparent, and technology-enabled governance. Delivering AI that is trusted is not an isolated process—it needs the collective effort of the entire organization. To achieve it, business leaders must consider three key questions: when to enact governance mechanisms, who is accountable for them, and how to operationalize governance and enable the organization.

In working directly with client organizations seeking to accelerate their adoption of AI, we created a framework to outline the capabilities necessary for good governance.

Capitalizing on the AI opportunity

From incremental improvement to complete reinvention, both established players and breakthrough entrants in myriad industries are seeking to capitalize on the potential of AI to cut costs and fuel innovation.

Properly applied, AI can have a material impact across the following facets of an organization's activities:

SCULPTING LEANER, FASTER OPERATIONS
AI can help improve efficiency and reduce costs.

PROVIDING TAILORED PRODUCTS AND ADVICE
AI can facilitate the personalization of services while maintaining scalability.

CREATING UBIQUITOUS PRESENCE
AI can help get products and services to customers how, when, and where they are needed.

DRIVING SMARTER DECISION-MAKING
AI can help process large volumes of data to deliver better business insights.

DISCOVERING NEW VALUE PROPOSITIONS
AI can help come up with new offerings and ways of working.

Despite the tremendous opportunities AI presents, however, many organizations and industries have been slower than expected in unlocking its potential. We explored the reasons for this adoption lag in our report *Canada's AI imperative*, but suffice to say that trust is a major factor. Uncertain regulatory environments, data security and privacy, and reputational damage all present risk.

Fortunately, there is an approach that lets organizations move forward on their AI agenda with confidence. The first step is to address the risks.

Addressing the risks of AI

Because AI is new and unfamiliar, and because the consequences of mistakes can be considerable, organizations are apprehensive about adopting AI. To manage the associated strategic, financial, and reputational business risks, it's important to first understand the inherent risks that exist in the design, development, deployment, and maintenance of AI systems.

Understanding AI risk

Business risk

Business risk broadly encompasses the strategic, financial, and reputational risks that an organization can accrue by using or developing AI.

Sources of risk

Organizations will have to manage four predominant sources of AI risk:

Data
... including the details of its collection, use, and exchange

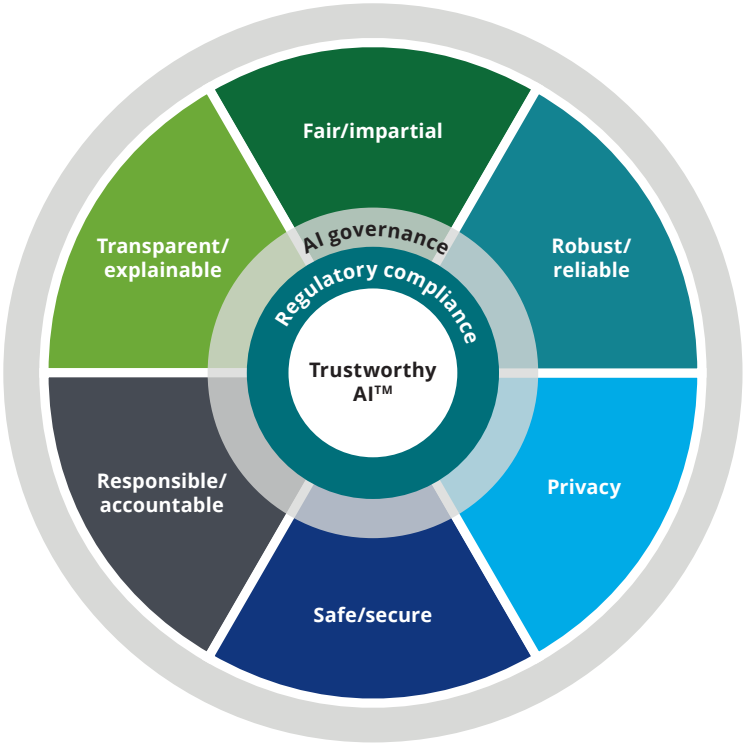
Models
... that derive predictions and insights from data

Technology
... and processes that form the complete AI system

Interaction
... between people and AI systems in making decisions and taking action

These sources of risk are where organizations can make strides in building mitigation strategies and policies. The prudent selection of data, the determination of how it can be used, and the governance decisions made during AI model development are examples of how organizations can mitigate risk and build trusted AI systems.

Deloitte's Trustworthy AI™ Framework



Regulatory compliance

Organizations face uncertainty in the regulatory landscape. Regulations to govern AI are at different stages of maturity and vary based on jurisdiction and industry. As regulators race to define the scope, applicability, and enforceability of regulations, the legal landscape can also change quickly. Organizations will need to monitor developments and be prepared to adapt quickly to meet new guidance. AI regulations are expected to have broad coverage over the inherent-risk areas of AI.

Inherent AI risk

Deloitte's Trustworthy AI™ framework has six components, which we will frame here as risks to be identified and mitigated. In our experience working with clients, we have found two complementary risk areas that both overlap with and can be addressed separately from the six components.

Fair/impartial	**Robust/reliable**
Organizations have a responsibility to ensure their AI systems do not create or perpetuate bias, and that groups are treated in a way the organization would consider to be fair.	Organizations must ensure their AI systems produce consistent and reliable outputs, performing tasks (and sometimes failing) as expected.
Privacy	**Safe/secure**
Organizations must ensure their AI systems are developed and deployed in consideration of an individual's consent and privacy rights, and that they can effectively protect personal information.	Organizations must thoroughly consider and address external, physical, and digital risks, among others, and communicate those risks to users.
Responsible/accountable	**Transparent/explainable**
Organizations must clearly articulate the ongoing roles and responsibilities of individuals, groups, and functions in the trustworthiness of an AI system.	Organizations must understand, interpret, and, in manyw cases, communicate how data is being used and how AI systems make decisions.

Complementary risk areas

Acceptable use	**Third-party liability**
Organizations must consistently assess the intended and unintended consequences of their AI systems and evaluate their alignment with organizational and societal values.	Organizations that rely on third parties for data, system development, deployment, or maintenance have a responsibility to hold these third parties to the same trusted AI standards the organizations observe.

Updating risk management processes

Addressing AI risk effectively will require the evolution of existing risk mechanisms and the creation of new governance processes dedicated to delivering AI that is trusted.

Existing risk mechanisms and processes spanning technology, privacy, cyber, compliance, etc., should be updated to reflect the new means by which AI systems can introduce risk. For example, a third-party risk management framework might be updated to reflect the potential roles a third party plays within an AI system; namely, as data providers, as model developers, as model owners, and as computing infrastructure providers.

New risk management processes must be created to address risks an organization might not have managed before, such as explainability or acceptable use. Organizations must build new capabilities to identify, mitigate, and manage these.

Taking a life-cycle approach to AI governance

While executives may understand the importance of addressing the risks associated with AI by creating new risk management practices and updating old ones, it's also vital they don't overlook the importance of choosing the right time to act.

Key questions to consider while determining when to enable governance include:

When should risks inherent to an AI system be identified?

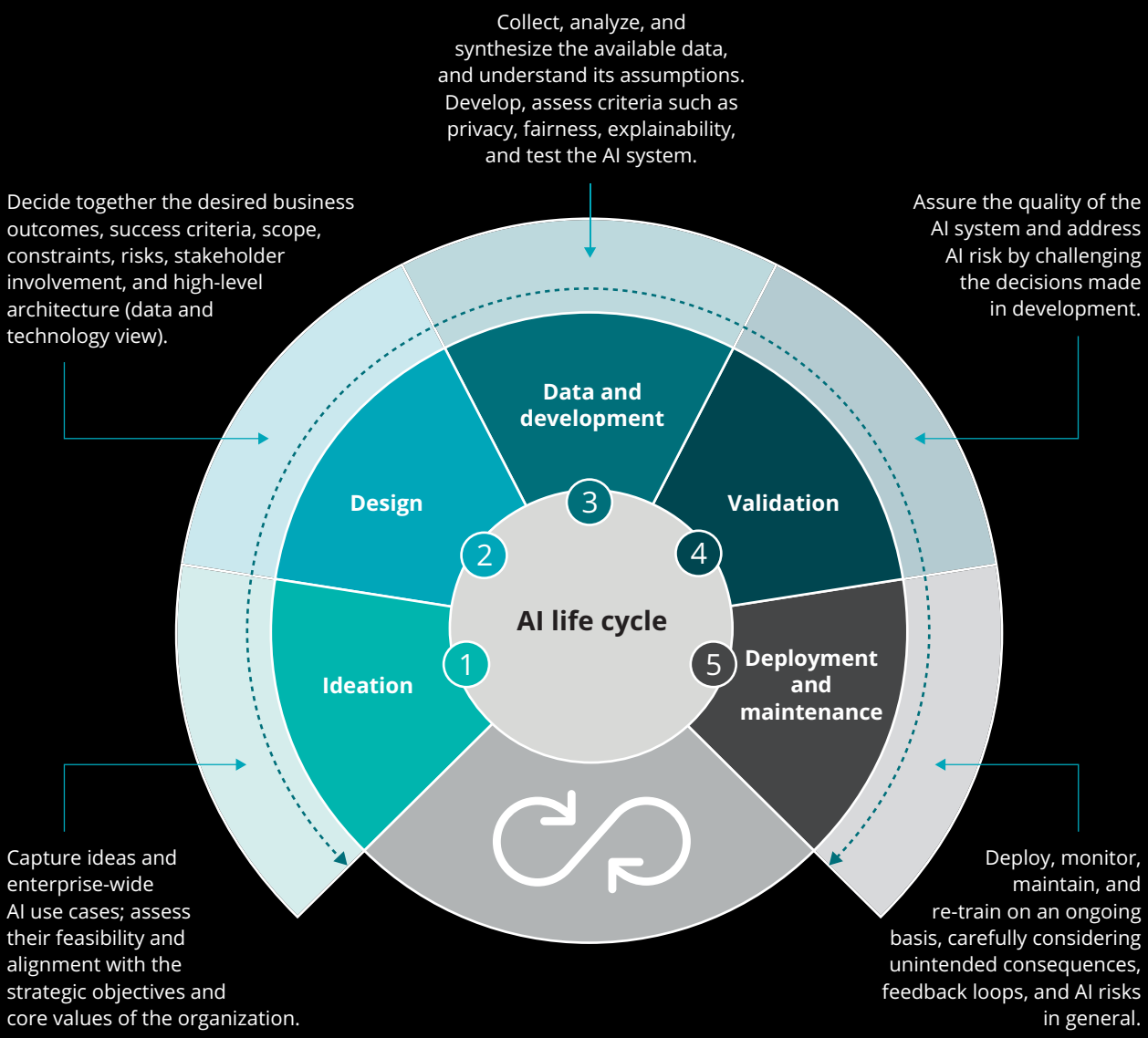
When should risks be addressed?

When should AI governance activities (e.g., approvals, tollgates, controls) be enacted?

When do key accountable parties and other stakeholders get involved?

Answering these questions at the right point in the AI life cycle can help organizations ensure that they are able to identify risks and make informed design decisions.

The **AI life cycle** helps organizations understand and decide the best timing for governance mechanisms. It represents the distinguishable, standardized stages of AI system development. Here's an overview:



While the life cycle is depicted sequentially, organizations will often iterate, especially through the inner three stages. Several AI governance capabilities should be designed in the context of the AI life cycle. Others affect the organization more broadly and are not distinguished by individual AI systems. Our framework for operationalizing trust in AI highlights this relationship and expands upon the capabilities required for robust AI governance.

Operationalizing trust in AI

Corporate strategies, and industry standards and regulations:
Ensure the organization's corporate strategy incorporates AI development and deployment priorities, and that it considers industry standards and applicable regulations.

Internal standards and policies:
Ensure the incremental risks of AI and mitigation strategies are reflected within existing internal standards and policies (privacy, security, vendor management, etc.).

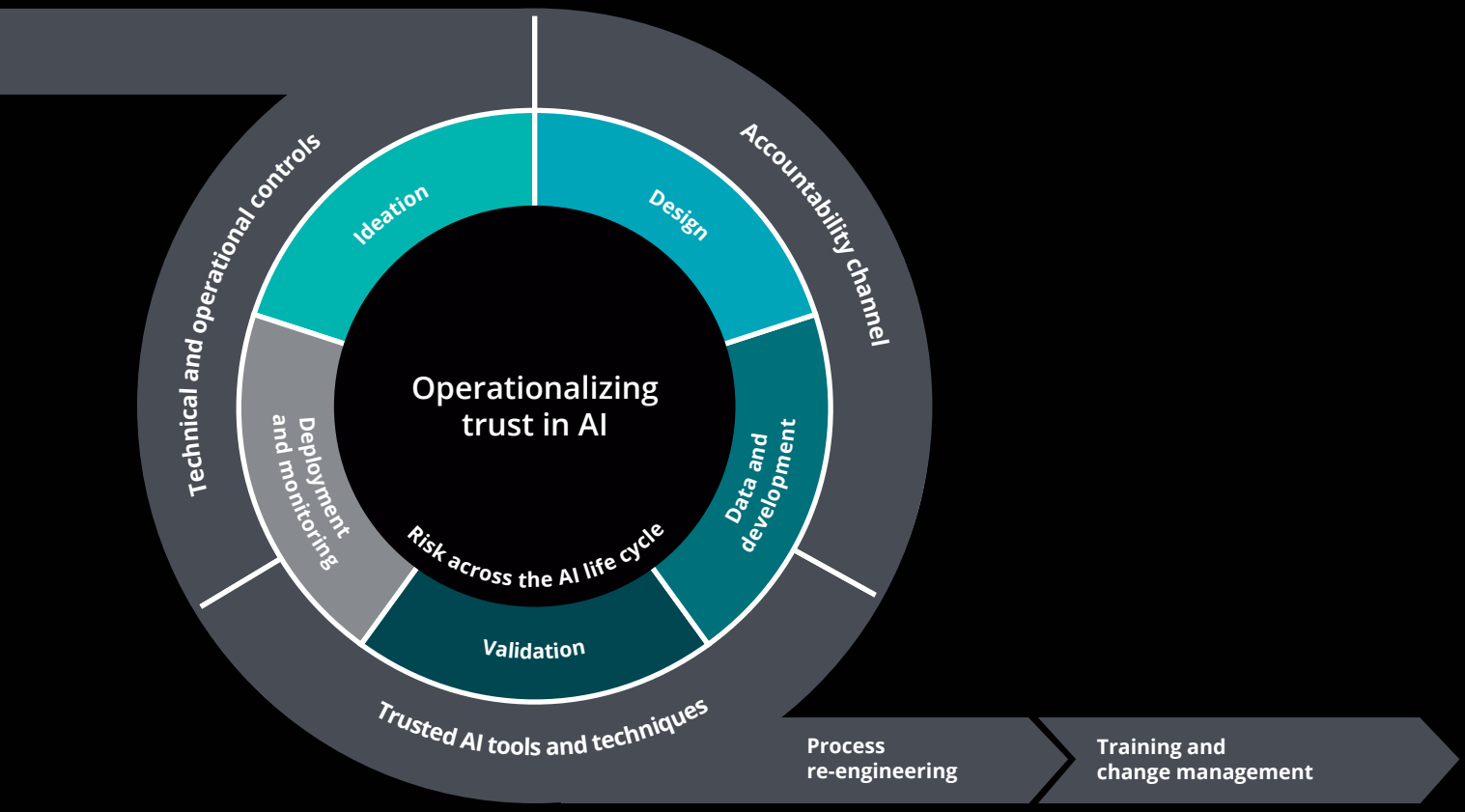
Controls:
Establish technical guardrails in the design of AI solutions to prevent specific actions from being completed.

Process re-engineering:
Place humans in the loop at critical points where the AI risks to consumers, employees, regulators, and the bottom line are unacceptably high.

Accountability:
Hold appropriate teams across the organization accountable for decisions on the selection, development, and deployment of AI use cases and systems, and empower them to remediate risks.

Tools and techniques:
Establish tools to dynamically monitor risks in AI systems. Integrate risk-mitigating and trust-building tools and techniques into the delivery of enterprise AI applications.

Training and change management:
Ensure customers, employees, shareholders, and other stakeholders are made and kept aware of the organization's perspectives and actions as they relate to the applications of AI.



Beyond the highlighted trusted AI governance capabilities, organizations will require two foundational capabilities for effective operationalization:

Enterprise data governance

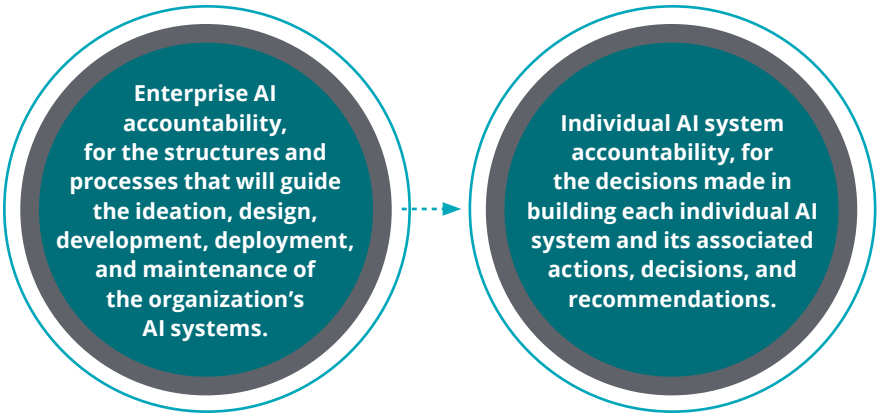
Effective enterprise data governance practices are critical to ensure AI delivers the required benefits and is aligned to business strategy. Regardless of the maturity level of their data capabilities, organizations should invest in a sound enterprise data governance program.

Enterprise risk management

AI governance efforts and enterprise risk management practices need to align and be tightly integrated for effective operationalization and adoption. Existing practices, principles, and processes should be employed and enhanced to address AI-associated risks across the enterprise.

Determining who is accountable for delivering trustworthy AI

It's important to address accountability early in an organization's AI governance journey. The deliver of trusted AI systems relies on two chains of accountability:



When an organization has implemented robust enterprise AI accountability, individual AI system accountability is largely ensured. A number of strategies—such as forums, teams, processes—can be used to deliver strong enterprise AI accountability. Two in particular are important ones: an oversight committee and a centre of excellence.

AI oversight committee

Executives and board directors must ask themselves whether they have the right forum to make decisions about AI development and deployment. They must also consider if there are clear accountability chains for senior leaders and whether the right people are represented in AI governance. To this end, it's imperative to establish a trusted AI committee comprised of representatives from the lines of business, technology, risk management, and other critical groups and functions (e.g., legal, regulatory, privacy, ethics) as well as AI subject matter experts. Its mandate should be to uphold the organization's commitments to building trusted AI while establishing robust AI governance and during its effective application.

AI centre of excellence

Most organizations adopt a federated, centre of excellence (CoE) model for the delivery of AI systems, and have designed and staffed them to support various groups or functions. They supply their own technical subject matter expertise to accelerate AI adoption within these areas. Many have tasked their AI CoE with ensuring effective AI awareness and training across the organization. However, leaders should also consider that their AI centre of excellence may also be suited to:

- Enabling, governing, and guiding the enterprise on responsible AI systems development and deployment; that is, taking on some of the oversight activities that build trust in AI
- Enabling the deployment of responsible AI solutions, which includes building, acquiring, and managing the tools the organization will need to build trusted AI systems
- Distributing enterprise-wide reporting and insights (e.g., on AI adoption, AI inventory)

Independent challenge

Some organizations, including those in the financial services industry, have well-established model validation teams and model risk-management frameworks aligned to regulatory guidance in their respective industries and jurisdictions. These teams provide independent challenge and oversight to ensure the quality and fidelity of models.

For organizations with these teams and frameworks, key questions include:

- To what extent should the existing model risk-management and model-validation team capabilities be extended to include machine learning and other AI technologies?
- Should current model-validation scope be revisited to ensure appropriate vetting of AI models that include embedded tools, or appropriate vetting of embedded tools?
- Which subsets of AI systems need to follow the highest validation rigour (i.e., highly structured, independent validation)? How might the remaining AI systems be validated (e.g., peer validation, process/decision review, self-assessment)?
- Should the existing teams remain focused on vetting the reliability and robustness of the system or extend their scope to the other inherent AI risk areas (e.g. fairness—impartiality)?

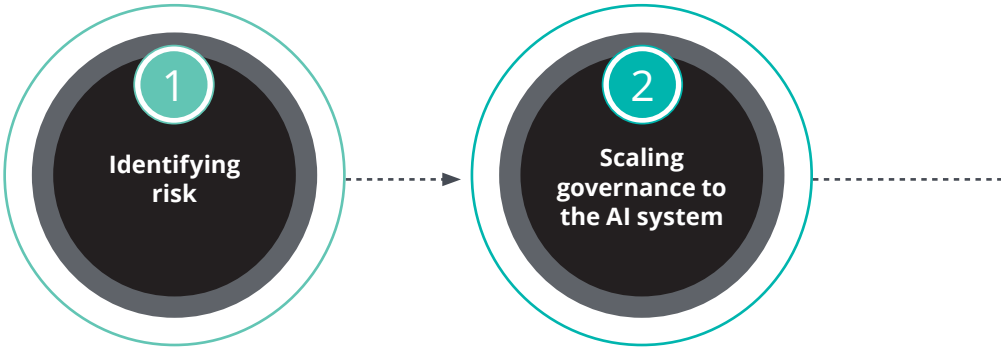
Organizations without independent challenge functions must consider how AI systems can be appropriately vetted prior to their deployment. They should consider the extent to which the validators are independent from the AI system developers.

Deloitte's framework elaborates on what makes an effective AI accountability model, including governance structures, decision-making forums, and the enhanced responsibilities of the AI CoE. It also sheds light on the involvement of different stakeholders in the development and deployment of AI systems.

Strategies to operationalize governance

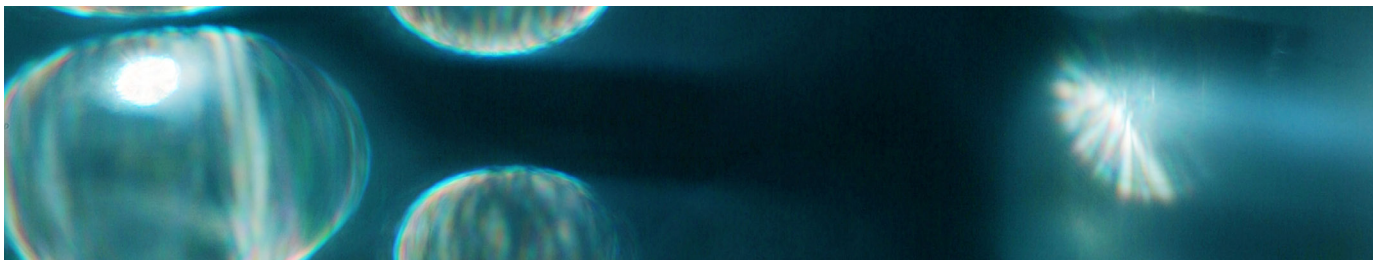
Effective, trusted AI governance is operationalized across three dimensions: people, process, and technology. Having discussed the importance of accountability (who is involved), let's turn our attention to the processes and technology that enable the development and deployment of trusted AI systems: the how.

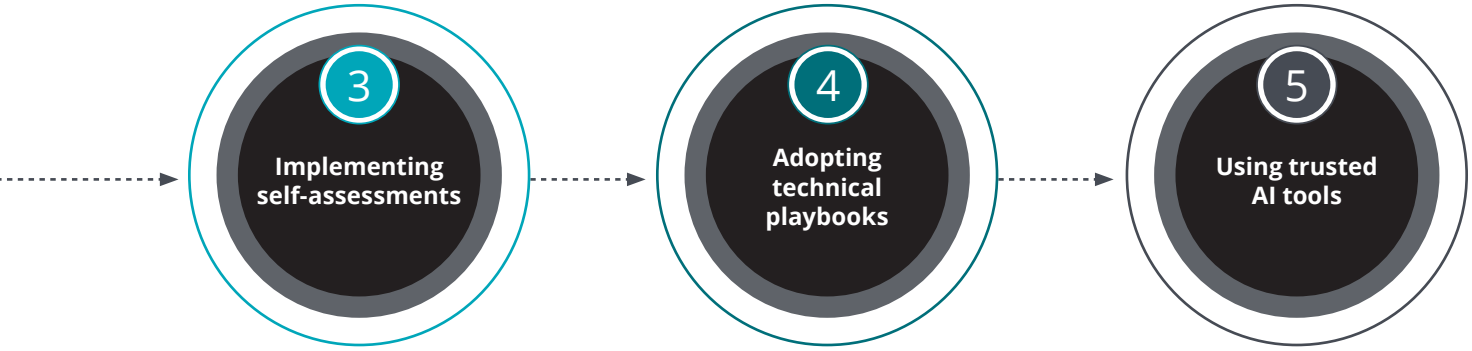
Organizations must strike a balance between accelerating the use of AI and having the right AI governance in place to ensure trustworthiness. There are five critical strategies:



Organizations must identify potential AI system risks early in the AI life cycle. This will allow AI system owners and developers to make the right design, development, and deployment decisions to build trust. Doing so will also limit the amount of redevelopment work. A common pitfall when risks are assessed and mitigated at the completion of the development or validation (quality-assurance) stage is necessary rework, or even losing time and financial investment if the project cannot be amended and therefore cannot be implemented.

Addressing AI risks involves a complex ecosystem of stakeholders who provide guidance and mitigation strategies, based on their functional expertise. Privacy SMEs, legal and compliance SMEs, security SMEs, senior data scientists, ethicists, and cross-functional groups of business leaders each play a role in addressing AI risks. An AI system owner should be aware of which stakeholder groups need to be consulted in order to be successful, and by extension, which stakeholder groups may not be necessary given the attributes of the AI system. This demands a more nuanced approach than standard risk tiers, which can be facilitated by understanding the key parameters and attributes of an AI system.





Organizations must rely heavily on AI system owners and developers to identify risk and scale governance. To extract the right amount of information, Deloitte has developed a Trusted AI self-assessment. Among its features, the tool collects key parameters of the AI system to gauge which AI risks require further attention and analysis, and then:

- Deliver actionable guidance to AI system owners and developers to inform their design and development decisions
- Triage AI system owners toward the SMEs (groups/functions) they must engage with to manage these risks

The tool also supports AI system inventory and aggregate reporting, one key to an organization's understating of its AI adoption, ROI, aggregate risk profile, and other relevant insights.

Technical playbooks provide developers building AI systems with tactical, situation-specific guidance. Their scope focuses on the types of AI systems most common to the organization. Playbooks demonstrate how certain techniques are applied, and contain references to open source or procured tools and resources. They also contain techniques that have been tested and are expected to have an extended shelf life, though they still need to be revisited periodically to update the techniques, examples, and references.

Fairness and explainability are risks well-suited to technical playbooks because their mitigation is applied at the individual AI system level and often require technical (programmed) methods.

Deloitte has created both fairness and explainability playbooks that focus on the types of AI systems we have seen our clients developing and deploying. Although not replacements for a robust training curriculum, they are critical resources for AI system developers.

After AI system owners and developers have used a self-assessment to identify inherent AI risks and make informed design decisions, and after they consulted technical playbooks to understand at a granular level what actions to take, they will require software tools to improve upon their AI systems. These tools can be open source or acquired solutions and are designed to address risk areas like fairness, explainability, and robustness. As organizations weigh the costs and benefits of building vs. buying solutions they will require a comprehensive understanding of the software landscape including real costs, customization, and the levels or proficiency required to effectively leverage the tools.

Deloitte has both analyzed the trusted AI tool landscape and has built trusted AI tools addressing key AI risk areas.

Change management to operationalize governance

As organizations embrace AI, they understand that it can bring significant transformational change to the workforce. As such, they must adopt change management strategies to apply a structured approach to transitioning to the desired future state. Companies should include their perspective on trusted AI in their communications and include AI governance considerations in employee onboarding and training.

Communications

A comprehensive strategy for communicating an organization's policies and strategies for mitigating the risks of AI will be crucial for demonstrating commitment to responsible AI practices.

Communications should be tailored to different audiences, both external and internal.

- ***To external stakeholders like users, customers, shareholders, and the public:*** Organizations must be clear about their dedication to using AI responsibly. This includes being transparent about what principles or policies they will follow.
- ***To internal stakeholders like employees, senior leadership, and board members:*** Communications should be clear and include details on the internal processes and procedures. Internal stakeholders need to understand how they will be involved and what their responsibility is in mitigating risk.

Education and training

In addition to clear communications, education and training can be delivered to various stakeholder groups to prepare them to contribute to the organization's responsible AI goals.

Executives and board members

- Leaders must be engaged early and often, and be provided with the necessary resources to be able to understand and contribute to strategic decision-making on mitigating both inherent AI risks and business risks.
- Boards must understand the risks that AI poses to the organization and the governance activities that can be employed to mitigate those risks. They should be empowered to know what questions to ask about AI.

Employees building AI systems

- Education for developers must focus on building employee awareness of inherent AI risk and potential business risks. Employees must be empowered to use the organization's existing technical playbooks to ensure systems are developed in alignment with the organization's responsible AI commitment.

Employees working directly with AI systems

- As roles and responsibilities shift as AI is adopted, employees working directly with the systems must be equipped with the proper training and tools to address emerging expectations.
- Education should focus on the awareness of AI system risks, limitations, and assumptions to help staff make better AI decisions. This includes training certain employees to understand and be able to execute mitigation protocols when and where necessary.

People leaders

- Managers and team leaders must understand how AI systems and AI risk could affect their teams and be able to help their teams flourish.
- Managers and team leaders should also have strong knowledge of governance structures and accountability chains, and be able to effectively triage risk incidents through the proper channels.

Broader organization

- Any AI awareness training that is presented to broad audiences within the organization should be complemented by education on AI risks and the organization's commitment to anticipating and mitigating them.

To prepare clients embarking on AI transformation journeys, we deliver curriculum-based training sessions on AI risks and governance tailored to different audiences.

These interactive sessions are offered through the Deloitte AI Academy and can be delivered in person and online. The academies are suitable for senior leaders, business and risk leaders involved in AI system development, and practitioners who are executing business and technical initiatives.

Contact

Preeti Shivpuri
Data and AI Governance Leader
Omnia AI
pshivpuri@deloitte.ca

Contributors

Nira Sivakumar
AI Strategy Leader, Omnia AI

Michael Vinelli
Manager, Omnia AI

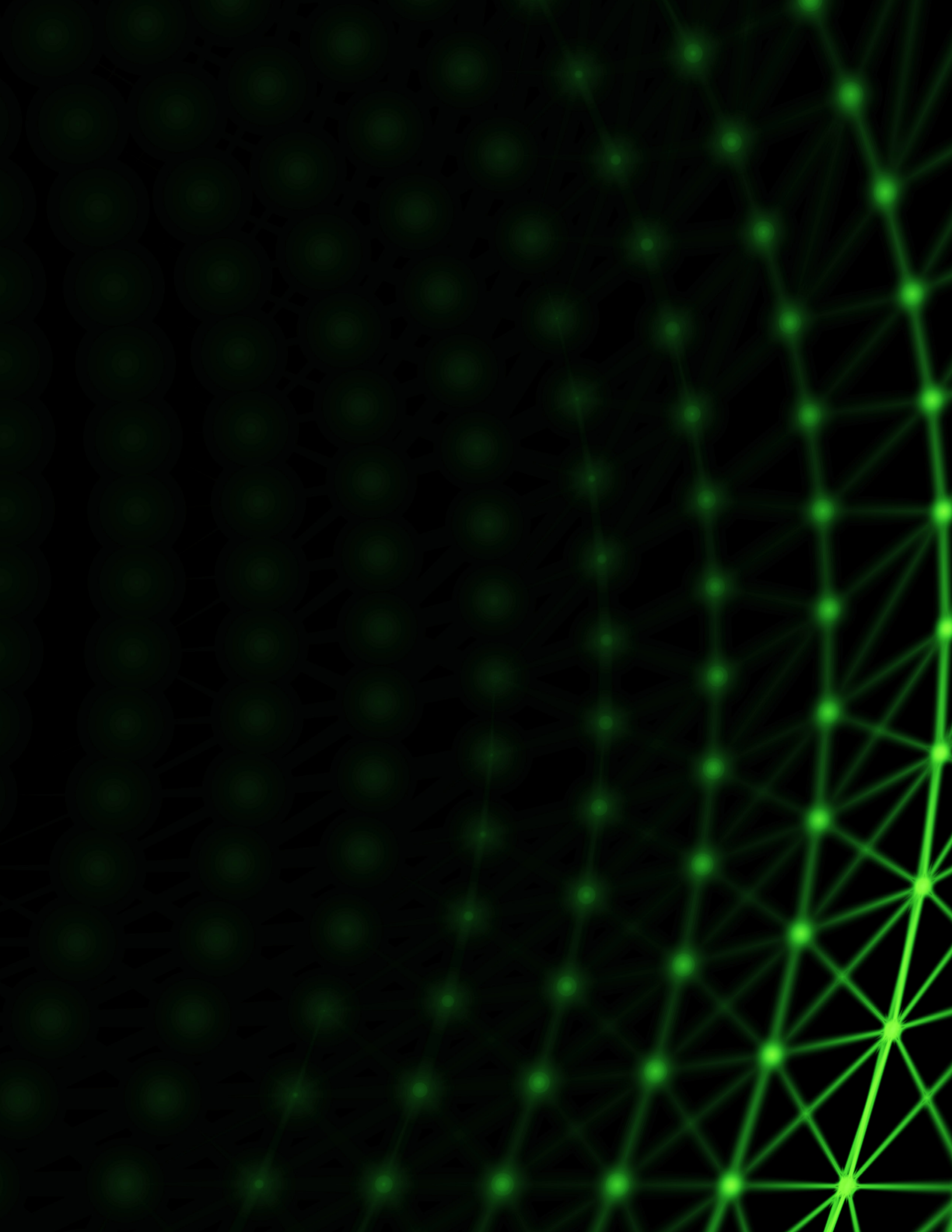
Amandeep Singh
Manager, Omnia AI

Monika Viktorova
Consultant, Omnia AI

Acknowledgements

Ishani Majumdar
Manager, Financial Services Consulting

Denizhan Uykur
Senior Consultant, Monitor Deloitte



www.deloitte.ca

Designed and produced by the Agency | Deloitte Canada. 20-3287597T