Adopting a holistic end-to-end Responsible AI strategy

Why and how the private sector should proactively establish AI governance

EY

Shape the future with confidence

The better the question. The better the answer.
The better the world works.

Introduction

Corporate adoption of artificial intelligence (AI) initiatives has increased drastically due to the recent generative AI wave. At the same time, global AI regulations are continuing to pass and will become enforced for various regional jurisdictions, industries and private sector's overtime. Academia and public sector guidance are falling behind private sector adoption and expansion of AI usage; therefore, businesses (particularly those in high-risk sectors such as health care and finance) should proactively establish Responsible AI policies and AI Governance procedures to help ensure corporate adoption and social responsibility when using and solutioning AI systems with customer and corporate data.

Responsible AI is sector-agnostic, and we are seeing an increase in demand for AI Governance across clients. There are unique risks associated with AI, even for the seemingly lowest risk sectors. Even if your corporation is in a field that does not commonly collect or utilize sensitive human data (i.e., PII, PHI), there are still various types of AI risks that can negatively impact your business in the long run if a responsible AI strategy is not adopted.

EY professionals have developed AI Governance frameworks, policies, processes and procedures to empower companies to get ahead of the curve by adopting Responsible AI strategies for safeguarding against AI bias, potential harms, reputational impact and compliance with evolving regulations.

"The why" — importance of Responsible AI

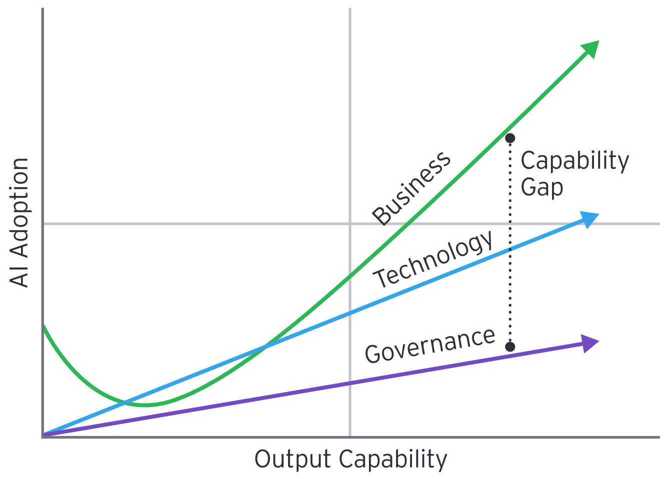
Over recent years as rapid AI adoption scaled into generative AI (GenAI) solutions, there has been increased scrutiny by the public on how corporations are designing and training AI with the potential to become biased against humans. Risks are increased particularly when using pre-trained generative AI models where foundational training data and AI model decision-making are not easily transparent and could lead to unintended harm to corporations or humans.

Academia and legislative bodies around the world are introducing guidelines for ethical and Responsible AI; however, the pace at which these guidelines are rolling out and being adopted are not catching up to the pace at which US private sector businesses are innovating and releasing AI products to global consumers. By 2026, Gartner predicts 50% of governments worldwide will enforce use of responsible AI through regulations, policies and the need for data privacy.[1]

To protect against potential damages to an organization, it is no longer enough for a company to publicly adopt and value Responsible AI principles. It is important to help ensure companies take action, guided by the principles, within the organization by setting up AI Governance bodies and incorporating new policies and procedures within data science organizations to begin adopting responsible ways of designing, producing and maintaining AI systems. It is equally important to educate the people who are funding, designing, building and delivering AI solutions about the tangible actions that should be adopted in order to transform the organization into a responsible yet innovative corporation.

The rate of growth of AI is far outpacing the regulations and regulatory bodies across the world are taking notice.

Confident adoption at speed and scale remains the big challenge



- Market growth will accelerate
- AI can be the true game-changer
- Growth in AI adoption and technological advances is only one part of a bigger story

[1] "AI Regulations to Drive Responsible AI Initiatives," Gartner, https://www.gartner.com/en/newsroom/press-releases/2024-02-29-ai-regulations-to-drive-responsible-ai-initiatives, 29 February 2024.

“The how” — main pillars to Responsible AI

AI is not new, and it’s very likely that organizations already have some level of Responsible AI practices integrated into their existing AI processes, often with a focus on AI ethics within data science teams. However, an organization’s AI landscape includes a wider range of business functional and technical components that should have Responsible AI activities and policies incorporated.

Every company is unique in how their IT organizations are structured and level of AI operations maturity. Based on Forrester Predictions 2025, 40% of highly regulated enterprises will combine data and AI governance. The complexity of AI governance, already intense due to rapid technological innovation and the absence of universal templates, standards, or certifications, is set to increase further.[2] Leading practice for understanding how an organization can expand Responsible AI begins with evaluating the existing functions around AI portfolio management, IT Governance, AI ethics, AI risks, legal, privacy, and security. At a high level, these functions should already exist for traditional software development and data analytics. As companies continue to adopt advanced analytics, machine learning, and GenAI, updates to corporate governance policies and procedures should be enhanced with the EY pillars of Responsible AI strategy.

Pillars of Responsible AI

Pillar 1

Scalable GenAI Governance structure

Pillar 2

AI portfolio intake and risk evaluation

Pillar 3

Solution development lifecycle embedded Responsible AI

Pillar 4

Proactive monitoring and controls

[2] *“Predictions 2025: An AI Reality Check Paves the Path for Long-Term Success,” Forrester, https://www.forrester.com/blogs/predictions-2025-artificial-intelligence/, 22 October 2024.*

Pillar 1

Scalable GenAI Governance structure

Adapting existing IT or AI governance structures to encompass newer AI risks on top of existing data privacy and traditional IT risk policies will establish a good foundation for Responsible AI. Sixty-seven percent of business leaders say more work is needed to address the social, ethical and criminal risks inherent in the new AI-fueled future.[3] Defining clear roles and responsibilities for evaluating the company's risk profile based on industry sector will begin informing the focused AI-related risks to incorporate into Responsible AI risk assessments and policies.

A robust structure for AI Governance is derived of three lines of defense:

- **First line** – technical teams who actively manage risks through designing, building, deploying and monitoring AI solutions
- **Second line** – functional teams with oversight into AI portfolio and AI leadership decisions identify emerging risks through review and assessments of AI solutions
- **Third line** – IT risk and audit teams independent from AI operations hold the first and second lines of defense accountable for complying with government regulations, IT and corporate Responsible AI policies

Due to the ever-evolving landscape of risks associated with AI, companies should now consider creating new responsible AI roles within their first and second lines of defense teams with a core focus on driving Responsible AI activities. These proposed Responsible AI roles would require a specific skill set across technology, business, risk and compliance.

[3] *EY Pulse Survey of CEOs*

Pillar 2

AI portfolio intake and risk evaluation

Establishing a comprehensive AI portfolio management intake process will help mitigate the chance of negative impacts to the business caused by high-risk AI solutions. AI portfolio management and governance teams should employ initial use case selection frameworks and AI Risk Tiering assessments at the onset of a new AI idea. When a business function within the company proposes a beneficial case for utilizing AI as a process improvement or business intelligence aid, each idea should funnel through a formal AI portfolio intake evaluation.

The evaluation will encompass information that enable AI stakeholders to determine the validity of a use case across technical complexity (i.e., Is existing technology and data ready to support the solution?), business value (i.e., How will this solution impact end users and revenue growth, and how can we measure benefits?), and Risks (i.e., What are the potential ethics, compliance, regulatory, privacy and security risks associated with this use case?)

If a use case moves forward with approval and the project is funded, the next step is to classify the use case into an AI risk tier. EY risk tiering consists of several dimensions of risk domain assessments to determine how risky a given use case will be once live in production.

Ethical, legal and privacy risks: Accounts for data privacy, fairness, bias and regulatory risks by assessing the use case's utilization of AI model type, corporate intellectual property, and sensitive personal data

Data, algorithmic and development risks: Accounts for technical complexity by assessing risks based on input data, tech stack, model utilized, model output and development methodologies

Business risks: Accounts for potential harms to the organization by assessing impact to revenue, customer experience, regulatory compliance and corporate reputation

The outcome of a use case's risk tier classification will determine two things:

1. Specific RAI controls and model monitors to be implemented with the AI system in production. Monitors can include tracking of performance accuracy, biases, end-user misuse, adversarial anomalies and more. RAI controls can include governance activities such as tracking use case compliance against data privacy, data retention, security, IT, and/or applicable legal policies
2. The ongoing frequency of AI system reviews and evaluation from various groups within the AI Governance structure's three lines of defense throughout the AI system's lifespan in production

Pillar 3

Solution development lifecycle embedded Responsible AI

Every phase in the AI system development lifecycle has potential risks that should have proper oversight. Developing with Responsible AI considerations across all projects will enable companies to produce inherently risk-adverse AI systems. There are various categories of risk that exist across the AI lifecycle:

- Use case initiation phase: design risks
- Data acquisition and preparation: data risks
- Model training, experimentation and validation: algorithmic risks
- Deployment and monitoring: performance risks

The EY Responsible AI framework guides development teams to check and analyze for the specific risks that could occur during each phase. Risk-mitigation activities should be incorporated as part of the standard solution development lifecycle to address the nuanced risks that arise for a specific business problem.

For example, if the data set required for training AI models is not readily available due to lack of digitization, disorganized data domains, or scattered ownership of data across various business units, etc., then analysis of how imbalanced a data set is before moving forward with model training is a critical part of mitigating data and algorithmic risks. It is also important to help ensure transparency into what kinds of data set was utilized for model training, as it informs the potential scenarios of bias and inaccuracy that can occur from the AI system's outputs. Depending on the business problem the AI system is intended to solve, biased outputs could be statistical, resulting in performance risks such as inaccurate insights given to the business. Social biases can also occur with AI systems that utilize human demographic data, such as those commonly employed in HR functions.

Pillar 4

Proactive monitoring and controls

As an AI program scales, risk mitigation monitors will need to be identified and streamlined. Specific monitors should be configured based on use case risks identified. Adhering to safe AI practices requires monitoring across several areas to account and adjust for potential risks, many of which have been heightened by the advent of GenAI. A selection of key risk categories are noted below:

Hallucination: Generation of outputs or conclusions by an AI system that are not grounded in its training data or input provided, leading to potentially incorrect, nonsensical or harmful responses.

- Deter the model from producing unfounded or imaginary content
- Set parameters to screen content outside of prescribed use case
- Create metrics to identify anomalies across various dimensions that might signal hallucination

Data Leakage: The unintentional exposure or sharing of sensitive or confidential data, either through the AI model's training data, predictive outputs or metadata, which may lead to privacy violations or security threats.

- Safeguard user confidentiality by vigilantly monitoring and controlling data output
- Prevent inadvertent revelation of sensitive information, fortifying user privacy and security

Prompt Injection: The intentional manipulation of the instruction or query given to an AI model with the aim to trick it into producing harmful, misleading or inappropriate responses, bypassing built-in safeguards.

- Guard against attempts to manipulate the model into bypassing its own safety protocols
- Provide a basis for refining the robustness of the model's safeguards, helping ensure it remains impervious to exploitation
- Helping ensure model functions around the prescribed use case parameters

Toxicity: Harmful or offensive content generated by an AI system, whether in response to a specific input or on its own, that could cause harm, distress or discomfort to individuals or groups.

- Proactively identify and mitigate "toxicity," defined as the generation of harmful, offensive or inappropriate content
- Systematic reinforcement of content moderation protocols
- Preemptively neutralize content that could undermine user wellbeing or violate platform guidelines

Automating the controls for the above categories of AI model risks can be done both proactively and retroactively. Proactive controls can include measures such as scanning and gating against toxic language, data leakage, prompt injection or setting thresholds for context variance from use case purpose. These risks can be monitored at both the input and the output level. Retroactive monitors can reveal past data and model performance to identify areas of improvement, reoccurring issues, and emerging risks. Automating responsible AI controls enhances the ability to provide continuous, efficient monitoring and can be done at the level of the AI platform architecture and in post launch data reviews.

The success of a Responsible AI program lies in finding the intersection between technical controls and functional processes. To develop responsible, scalable and successful AI systems, data science and technology teams must follow guidelines and regulations set up by AI Governance and business functional bodies. Meanwhile, governance bodies must collaborate with technical teams to continuously refine these rules, based on emerging risks and changing societal context. Through this collaborative model, AI can be utilized in an ethical and beneficial manner, contributing positively to business strategies, customer trust and brand reputation.

Client case study

The client, an energy division of a large, multinational conglomerate, requested support in developing a comprehensive GenAI strategy to successfully move from a POC phase to a robust, comprehensive AI program. To develop their core strategy, the EY organization developed a Responsible AI Governance playbook to identify, mitigate and monitor for AI system risks. We partnered with the company as they began to develop their GenAI platform in order to proactively create processes for responsible development. Key outputs included Responsible AI lifecycle process flows, roles and responsibilities, risk tiering assessment, risk controls and monitoring enablers.

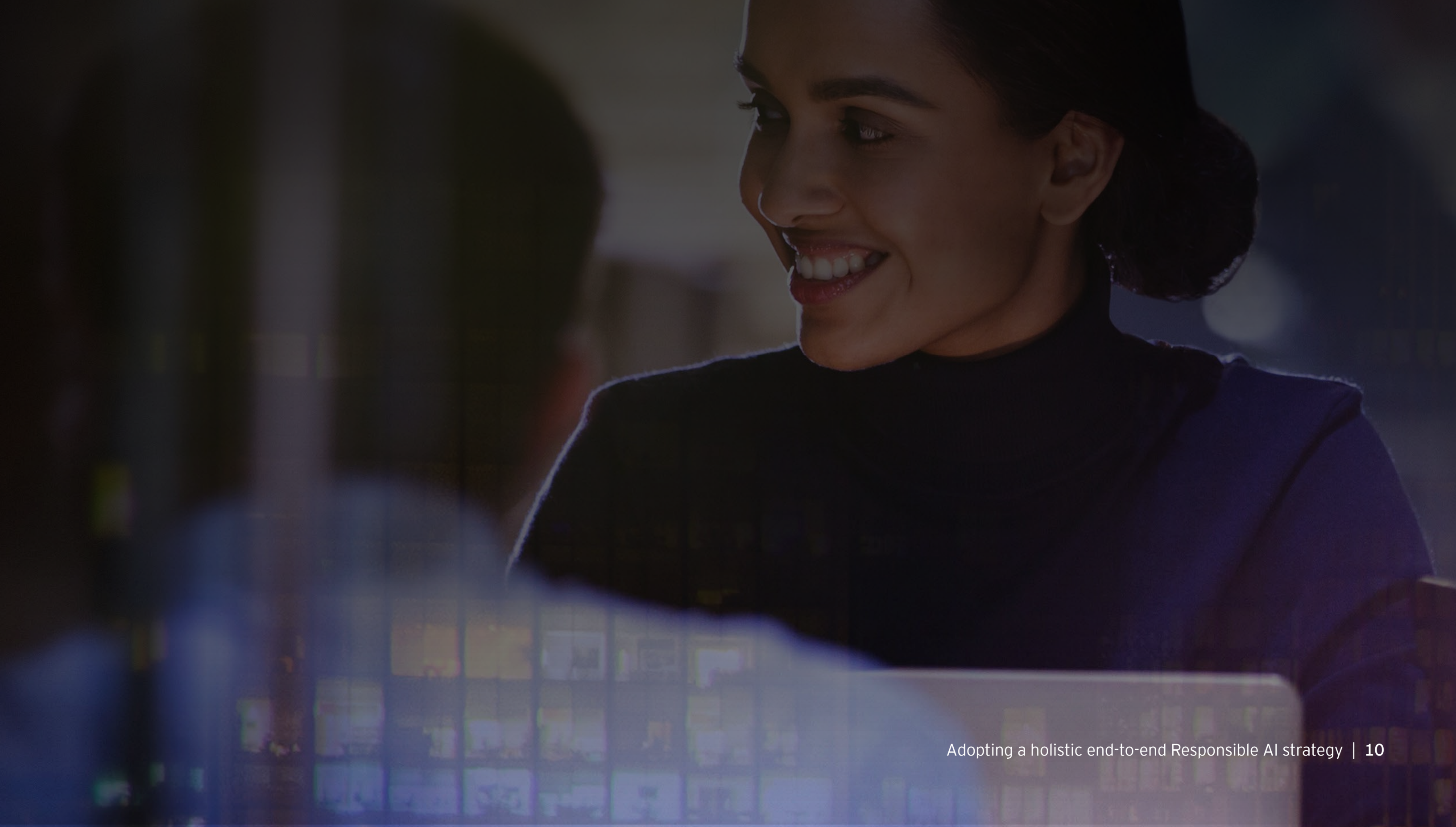
Operationalizing the new Responsible AI process, roles and responsibilities across the client's standard solution development lifecycle (SDLC) helped educate data science and engineering teams on the nuances of AI risks that can occur at every stage of the SDLC. Conducting Responsible AI procedures with a real use case helped development teams adopt a wider perspective for analyzing impacts the AI system could have on the broader organization and their own customers if risks are not identified and mitigated.

Summary

Responsible AI Governance is no longer optional. Maintaining a leading practice AI Governance model is fundamental to driving innovation and market competitiveness, and necessary for managing AI risk and maintaining regulatory compliance. Organizations should account for the below pillars to stay ahead of the curve of evolving AI risks, changing regulations and plan for the automation of risk management tools and controls to advance organizational structure along with new technology and ways of working.

1. An organizational structure that accounts for Responsible AI roles and responsibilities will help ensure adherence to safe, ethical principles and methods.
2. End-to-end portfolio management will enable oversight and management of risks as AI use cases grow in numbers and complexity. Standardized risks evaluations, specific to AI and contemporary regulations, will provide clear guidelines and measures to maintain compliance.
3. Responsible AI controls embedded in the software development lifecycle provide touch points and escalations paths to review and reassess risks identified in risk evaluations.
4. Monitoring and control capabilities and validation streamline review processes and protect against risks of future regulations and audits.

With the rapid technical advancements fueled by large language models, organizations need to stand up updated, robust operating models, roles, and processes to facilitate long-term AI program growth. Responsible AI will soon become a new global way of working, which opens the doors to new opportunities, skill sets, jobs and adoption of new methods of inclusivity and ethical AI.



Authors

Samta Kapoor
Partner
Technology Consulting
Ernst & Young LLP
samta.kapoor@ey.com

Mary Kryska
Partner
Technology Consulting
Ernst & Young LLP
mary.kryska@ey.com

William Smith
Senior Manager
Technology Consulting
Ernst & Young LLP
william.smith2@ey.com

Molly P. Donovan
Manager
Technology Consulting
Ernst & Young LLP
molly.donovan@ey.com

SCORE no.
2411-11095-CS
ED None

This material has been prepared for general informational purposes only and is not intended to be relied upon as accounting, tax, legal or other professional advice. Please refer to your advisors for specific advice.

ey.com