Centre for Information Policy Leadership

HUNTON ANDREWS KURTH

Building Accountable AI Programs:

Mapping Emerging Best Practices to the CIPL Accountability Framework

February 2024

Table of Contents

Table of Contents (continued)

Foreword

Artificial intelligence (AI) technologies have permeated nearly every aspect of everyday life, precipitating a transformative impact on individuals and society. While AI-powered tools can deliver a wide range of substantial benefits, they also carry significant risks. Thus, it is critical for organizations to implement robust, systematic AI governance frameworks and compliance programs with proper oversight, policies, procedures, tools, training, and reviews, that enable the effective weighing of risks and benefits and the adoption of appropriate risk mitigation measures. These programs are essential to creating public trust in the responsible development and deployment of powerful AI technologies and their beneficial impact.

For more than 20 years, the Centre for Information Policy Leadership (CIPL) has advocated for organizational accountability and a risk-based approach to regulation and compliance as the cornerstones of effective data privacy and wider data policy measures. Since 2018, CIPL has also promoted these as critical to the development and deployment of AI technologies and was an early contributor in identifying both challenges and practical solutions for AI governance and industry practices. This new CIPL report continues that work by showcasing how 20 leading organizations are developing AI governance and best practices on the ground. I am grateful to the organizations who participated in this project for sharing their trailblazing work in this fast-moving space.

Our research shows that organizational accountability is fundamental to the responsible development and deployment of AI. Organizations recognize the need to demonstrate AI accountability as a business imperative, especially as the expectations of consumers, business partners, shareholders, and regulators continue to grow. They also recognize that AI governance works best when it leverages knowledge from other disciplines, including data protection, information security, human rights, cybersecurity, and more.

Implementing and demonstrating accountability are continuous processes. Our paper's findings capture the early stages and emerging best practices in creating accountable AI programs. These best practices will continue to evolve as organizations test and audit their programs against developing technology, industry standards, and a growing body of regulatory and co-regulatory requirements. I am excited to see the next steps in this shared journey.

I do hope that the case studies and findings from our report influence other organizations to invest in digital accountability and build effective accountability measures into their AI programs that not only address the challenges of developing and deploying AI technologies responsibly, but also maximize their benefits. I also hope that our report builds consensus and shared expectations on what effective accountability in AI management looks like in practice. Further, I hope our report inspires lawmakers, policymakers, and regulators to create and promote incentives that will lead organizations to continue investing in accountable and responsible AI practices.

Bojana Bellamy

President
Centre for Information Policy Leadership

I. About this Report – Methodology and Objectives

Artificial Intelligence (AI) technologies are transforming and disrupting the world around us. They are impacting how companies operate and create new products and services; how governments make choices and deliver efficient and meaningful public services; how research organizations create new scientific breakthroughs; how data can be used for public good and public benefit; and how people around the world live, work, and connect. However, the benefits of transformative AI technologies also come with new and existing risks for individuals and society. Organizations developing and deploying these technologies must bear the responsibility of doing so in a trustworthy and accountable way that mitigates and manages the risk while preserving the benefits.

The Centre for Information Policy Leadership (CIPL) has advocated for, and provided thought leadership on, organizational accountability in data protection and broader digital and data policy for more than 20 years, including how organizations can demonstrate accountability through comprehensive data protection management programs.[1] In 2018, CIPL recognized that organizational accountability can serve as a key building block not only for effective data protection, but also responsible governance of AI, and began to focus on promoting demonstrable accountability in AI governance, including at the intersection of data protection laws and AI. CIPL has been advocating for the recognition of organizational accountability frameworks and programs as core elements of effective AI regulation, as well as the implementation of comprehensive accountability programs by organizations in the context of AI development and use.

Our new report on *Building Accountable AI Programs: Mapping Emerging Best Practices to the CIPL Accountability Framework* draws inspiration from CIPL's 2020 report on *What Good and Effective Data Privacy Accountability Looks Like: Mapping Organizations' Practices to the CIPL Accountability Framework*,[2] where CIPL collected and identified organizational best practices from leading companies' data privacy management programs and mapped these practices to CIPL's Accountability Framework. The report's purpose was to highlight organizational best practices being deployed on the ground and promote their wider adoption in the market. Our present report also builds on other CIPL work, including several CIPL white papers on the intersection of AI and data protection,[3] and CIPL's *Ten Recommendations for Global AI Regulation*.[4]

1 CIPL Organizational Accountability Project resources, available at https://www.informationpolicycentre.com/organizational-accountability.html

2 CIPL, What Good and Effective Data Privacy Accountability Looks Like: Mapping Organizations' Practices to the CIPL Accountability Framework, May 2020, available at https://www.informationpolicycentre.com/cipl-2020-accountability-mapping-report.html

3 CIPL, Artificial Intelligence and Data Protection in Tension, October 29, 2018, available at https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl-hunton_andrews_kurth_legal_note_-_how_gdpr_regulates_ai__12_march_2020_.pdf; Hard Issues and Practical Solutions, February 27, 2020, available at https://www.informationpolicycentre.com/ai-project.html; Artificial Intelligence and Data Protection: How the GDPR Regulates AI, March 2020, available at https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl-hunton_andrews_kurth_legal_note_-_how_gdpr_regulates_ai__12_march_2020_.pdf; CIPL's Response to the National Telecommunications and Information Administration (NTIA) Request for Comment on AI Accountability Policy, June 12, 2023, available at https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_response_to_ntia_ai_accountability_policy_june2023.pdf

4 CIPL White Paper on 10 Recommendations for Global AI Regulation, October 10, 2023, available at https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_ten_recommendations_global_ai_regulation_oct2023.pdf

Organizational accountability has long been established in law and regulatory guidance across a variety of corporate compliance areas beyond data protection, privacy, and AI. [5] Its origins can be traced back to certain US legal frameworks, such as the Foreign Corrupt Practices Act (1977) and the Sarbanes-Oxley Act (2002).[6] Organizational accountability is also recognized as a key building block of effective privacy and data protection regulation and compliance. A well-developed, comprehensive accountability framework or program provides organizations with the tools and processes needed to implement relevant legal requirements and standards, as well as internal ethics standards and other internal "best practice" goals.[7] In addition to compliance with applicable legal requirements and other objectives, accountability frameworks and programs can act as market differentiators by building trust from stakeholders, business partners, and regulators, and allow organizations to more effectively and broadly leverage data and AI technologies.

CIPL's Accountability Framework contains seven core elements of accountability: leadership and oversight; risk assessment; policies and procedures (including on fairness and ethics); transparency; training and awareness; monitoring and verification; and response and enforcement (Figure 1). This framework has been used by global businesses as a model for building comprehensive privacy and data governance programs, and it can equally be used in the AI context to build comprehensive governance programs that ensure responsible development and deployment of AI.[8]

Figure 1. CIPL Accountability Framework

Source: CIPL

The present report identifies and classifies key best practices and lessons learned from organizations with mature or rapidly advancing AI governance programs through the lens of CIPL's Accountability Framework, with the goal of making these best practices available to organizations that are just beginning to navigate the rapidly evolving AI landscape. To that end, CIPL identified 20 organizations of various sizes and from various industry sectors that have demonstrated a commitment to implementing accountable and ethical practices for the development and deployment of AI.

5 CIPL Discussion Paper on "Organizational Accountability in Light of FTC Consent Orders", November 13, 2019, available at https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_white_paper_-_organizational_accountability_in_light_of_ftc_consent_orders__13_november_2019_.pdf

6 CIPL White Paper on "The Concept of 'Organizational Accountability' - Existence in US Regulatory Compliance and its Relevance for a Federal Data Privacy Law", 3 July 2019, available at https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_white_paper_on_organizational_accountability_-_existence_in_us_regulatory_compliance_and_its_relevance_for_a_federal_data_privacy_law__3_july_2019_.pdf

7 CIPL, Q&A on Organizational Accountability in Data Protection, July 3, 2019, available at https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/%5Ben%5D_cipls_top_ten_recommendations_for_regulating_ai_in_brazil__4_october_2022_.pdf

8 CIPL Report of the CIPL Accountability Mapping Project, What Good and Effective Data Privacy Accountability Looks Like: Mapping Organizations' Practices to the CIPL Accountability Framework, May 2020, available at https://www.informationpolicycentre.com/cipl-2020-accountability-mapping-report.html; CIPL White Paper, Top Ten Recommendations for Regulating AI in Brazil, October 4, 2022, available at https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/%5Ben%5D_cipls_top_ten_recommendations_for_regulating_ai_in_brazil__4_october_2022_.pdf

Project participants included Accenture, AstraZeneca, BBB National Programs, Dun & Bradstreet, Cisco, eBay, Google, IBM, Mastercard, Meta, Microsoft, Onfido, PayPal, Prosus, RELX Group, SAP, Telefonica, and Vodafone, among others.

These organizations participated in extensive one-on-one interviews to document their respective AI governance practices. To supplement the interviews, CIPL reviewed relevant written materials regarding each organization's AI work that were either publicly available or provided directly by the organizations. The practices identified through this research were then classified, or "mapped," to the seven elements of the CIPL Accountability Framework.[9]

CIPL's project has the following objectives:

1. **Provide concrete empirical evidence** of best practices, including case studies, to show how accountability-based governance frameworks are demonstrable, enforceable, and effective;
2. **Share best practices and lessons learned** from leading global organizations in implementing robust AI governance measures that address all core elements of organizational accountability;
3. **Build global consensus** on how accountability can foster responsible development, deployment, and use of AI;
4. **Promote accountability** as an effective organization-wide business strategy that goes beyond legal compliance; and
5. **Inform the global debate** on AI regulation and AI governance oversight and enforcement.

9 See Appendix A for a comprehensive list of best practices in AI governance mapped to the CIPL Accountability Framework.

II. General Findings

CIPL's research revealed the following general findings and commonalities amongst project participants and accountable organizations:

1. **AI transformation, coupled with accountability in AI, is a top priority and business imperative.** Building confidence and trust from both internal and external stakeholders is essential to the successful adoption of transformational technologies like AI. Across all business functions, from engineering to the C-suite, organizations are striving to ensure that AI technologies are developed and deployed in a manner that enables users to reap the benefits of the technology, while preserving individuals' rights and societal interests and identifying and mitigating the potential risk of harms and noncompliance. Not only do external stakeholders expect organizations to act responsibly, but employees want to work for organizations that are developing and deploying technologies in an ethical and responsible way. Organizations have thus built accountable AI programs with buy-in from all functions and levels of the organization (e.g., by incorporating accountable governance AI into their guiding principles, by including AI-related topics and ethics into their mandatory corporate training, etc.).

2. **Accountable governance of AI is a smart business investment for long-term sustainable and competitive business.** By setting up their internal governance and programs now, accountable organizations hope to have a competitive edge as AI technologies become more universally adopted. They recognize that trust is an invaluable commodity. Therefore, though it may demand a large initial investment of time and resources, organizations believe that choosing to do the right thing now will be worthwhile and lead to greater success in the long term.

3. **"Tone from the top" is crucial for responsible AI programs.** Senior leadership and board-level executives are establishing and heralding company-wide principles, values, and commitments that enable accountable governance of AI, which are then cascaded and embedded throughout the organization. Organizations recognize that leading by example can ensure top-down awareness of ongoing efforts and foster a corporate culture that views accountability in AI governance as a shared, company-wide responsibility that every employee can and should contribute to.

4. **Guidance from regulators and policymakers is welcome as organizations prepare to implement emerging standards and regulations,** such as the EU AI Act, the US Executive Order 14110 on Safe, Secure, and Trustworthy Artificial Intelligence, the NIST AI Risk Management Framework (AI RMF), US Executive Order 13960 on Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government, or the ISO/IEC 42001:2023. Organizations are expecting forthcoming regulation of AI and are proactively preparing to comply with potential regulatory requirements by implementing new policies, updating their internal glossaries, and adopting AI-focused risk assessment frameworks. Organizations, especially those operating in the European market, anticipate that the EU AI Act will be influential globally and impact how other countries approach AI regulation.

5. **A risk-based, technology-agnostic approach is the most effective and appropriate approach for AI governance.** When managing and mitigating risks in AI development and deployment, organizations are assessing the specific use cases or applications of AI technology rather than the technology itself. Organizations believe that this approach is key to creating effective, future-proof governance structures that will adapt with the technology. Because AI technologies are advancing at a rapid pace, frameworks that are overly prescriptive could become quickly outdated and inhibit beneficial innovation and progress.

6. **Convergence around common AI-related terminology is urgent and essential.** Building consensus around AI-related terminology has been a challenge due to many factors, including differing priorities, scope, technical expertise, and contexts. Many are looking to the definitions set forth by upcoming regulations and standards, such as the EU AI Act, the NIST AI RMF, or the OECD AI Principles, to encourage convergence. To facilitate standardization within the organization, some have also created working glossaries for internal reference and use.

7. **Organizations are adapting and updating their governance frameworks to address new issues and risks, including those raised by generative AI (GenAI).** All organizations acknowledge that AI's rapid advancement, democratization, and widespread use pose novel questions concerning reliability, accuracy, copyright and intellectual property, misinformation and disinformation, bias and discrimination, and more. However, they agree that addressing AI-related issues should not require a complete overhaul of existing corporate compliance and governance structures but can be accomplished by adapting and updating existing ones. Furthermore, organizations believe that it is essential to take a risk-based approach and deploy timely, appropriate, and proportionate mitigation measures to secure the benefits of AI, address its relevant risks, and safeguard individuals' rights. Most organizations have a multi-pronged approach, adopting many of the following: promulgating policies and controls regarding internal AI development and use; offering AI training and education modules; creating additional oversight bodies, such as AI ethics committees; joining regulatory sandboxes; and incorporating red-teaming or adversarial testing for their AI models. Many organizations have also implemented GenAI specific measures, such as creating external-facing statements on how they will internally deploy GenAI tools and providing employees with GenAI-focused training.

8. **Multidisciplinary and diverse teams are the foundation for building and implementing accountable AI governance programs.** While there is no single "right" location for an AI governance team within an organization, many view data privacy teams as the logical points of departure for AI governance considering their expertise with data management programs, tools, and regulatory compliance. Regardless of where AI governance teams are situated, however, organizations agree that these teams must be cross-disciplinary and include representatives from relevant disciplines, including data science, privacy, legal, ethics, compliance, engineering, product, IT, and information security. Organizations are also upskilling their workforces by offering training courses and encouraging enrollment in external certifications and programs. Furthermore, organizations are hiring to increase diversity across a number of factors, including demographic (e.g., gender, age, religion, sexual orientation, and ethnicity), relevant expertise (e.g., technology, social, industry, legal, human resources, government relations, ethics), and geography.

9. **There is great value in building consensus on the appropriate elements of accountable AI governance and benchmarking against peers.** Whether through participation in large conferences or smaller benchmarking interviews, workshops, or roundtables, organizations stress the importance of engaging in informal peer benchmarking to help shape and guide AI accountability efforts. For example, a significant incentive for many organizations participating in the project was being able to participate in benchmarking workshops and share best

practices with other organizations, while hearing from CIPL what the current best practices and emerging trends are for accountable governance of AI. Project participants also agree that CIPL's Accountability Framework is a useful, well-established foundation for building an effective AI governance program that can remain adaptable and flexible to accommodate forthcoming standards and regulations.

10. **Accountable development and deployment of AI is a continuous journey and an iterative process.** All organizations stress that their AI governance programs and controls are continuously evolving; they are adapting and calibrating them to technological and regulatory developments, as well as to the knowledge and experiences they are collecting along the way. This captures the very essence of CIPL's approach to accountability and reflects our overarching principle: the higher the potential risk of harms, the more organizations must do to demonstrate the elements of organizational accountability.

III. Examples and Best Practices for Implementing the Core Elements of Accountability

Through this project, CIPL collected a wide range of best practices, examples, and case studies concerning specific measures and processes organizations use to create accountable AI programs, in the hopes that they may be instructive for organizations that are developing their own AI governance programs.

These practices have been sorted and categorized according to the seven elements of CIPL's Accountability Framework. As noted above, accountability in AI development and deployment is an iterative process. By organizing best practices and activities in a consistent framework, like the CIPL Accountability Framework, organizations can develop a repeatable and systematic approach to AI governance. We encourage organizations seeking to develop their own accountable AI programs to follow a similar methodology or framework, which will help them demonstrate accountability to both internal (e.g., boards of directors, the C-suite, executive committees, employees) and external (e.g., regulators and investors) stakeholders.

1. Leadership & Oversight

Leadership and oversight are essential to ensuring holistic and demonstrable accountability in an organization's AI governance program. This accountability element can be demonstrated through an organization's commitments to creating effective governance structures, appointing appropriate personnel to oversee them, and promoting awareness and support across all functions of the organization.

CIPL has identified a range of practices across organizations that reflect the elements of leadership and oversight. A few examples include:

- Establishing "tone from the top" and demonstrating a commitment to advance ethics, values, and specific principles in AI development, deployment, and use;
- Implementing systematic processes and escalation pathways for AI-related decision making;
- Establishing AI ethics oversight bodies or committees (internal or external) to review risky AI use cases and promote ongoing improvements to AI practices;
- Appointing a board member for AI oversight;
- Appointing a responsible AI lead, AI officer, or AI champion;
- Setting up an internal interdisciplinary AI board or AI committee;
- Establishing organization-wide AI ethics principles;
- Ensuring inclusion and diversity in AI model development and AI product teams;
- Creating a centralized governance framework with oversight from the top that still provides flexibility within internal teams;
- Expanding the remit of privacy teams to include AI-related responsibilities;

- Leveraging the expertise of other relevant teams (e.g., engineering, data science, legal, ethics and compliance, etc.) to ensure multidisciplinary, cross-functional AI teams; and
- Encouraging employee reporting throughout all levels of the organization by offering escalation pathways to resolve potential AI-related issues.

1.1 Commitment from the top ("tone from the top")

Organizations increasingly recognize that investing in responsible, accountable governance of AI is good for business. Digital trust and confidence are key to growing a sustainable and robust long-term business. C-suite executives and board members are having to meet the growing expectations of customers (both consumers and businesses), investors, regulators, and media by demonstrating their commitment to responsible and ethical digital business practices and by ensuring the responsible development and deployment of novel and transformative technologies, including AI.

For most of the organizations participating in CIPL's study, the need to invest in trust—and the decision to act on it—came from senior-level executives who had a comprehensive understanding of not only the organizations' business needs, but also the pressures and expectations they faced, both internally and externally. Senior leaders also frequently relied on significant input from employees in different roles and at different levels throughout the organization to help shape their approach to AI governance.

> **"At Cisco, our overall corporate purpose is to power an inclusive future for all. Privacy and Responsible AI are business imperatives core to furthering that purpose—on top of legal compliance and 'it's the right thing to do.'"**
>
> **Harvey Jang**
> **VP, Deputy General Counsel & Chief Privacy Officer, Cisco**

All organizations report a growing impetus to invest both monetary and human capital into developing responsible AI programs. This includes creating new roles and teams to specifically address issues raised by the development and deployment of AI. Many organizations attribute this increased investment to the rapid implementation of AI across all business sectors, the growing excitement surrounding the potential and trajectory of GenAI, and concerns about potential risks and harms. Many organizations also expanded the roles and remit of their data privacy programs and teams to include data ethics and digital trust issues and create synergy with their broader ESG (Environmental, Social, and Governance) strategy.

The role of corporate leadership in setting the "tone" and promoting a culture of accountability has become essential in successfully implementing a top-down approach. This approach first requires leaders to establish company values, expand or update existing codes of business conduct and ethics, and formalize commitments to responsible development and deployment of AI technologies, all in written documentation distributed both internally within the organization and externally to consumers or business partners. Leadership can then appropriately delegate responsibilities to ensure that the uptake of these directives and objectives, as well as the broader goal of responsible AI governance, becomes a shared responsibility across all business functions, geographies, and employees.

> **CASE STUDY 1: Periodic company-wide open meetings with the CEO**
>
> A digital communications technology organization conducts periodic company-wide meetings where the CEO will share senior leadership's views, objectives, and organizational direction with employees and create an open space for questions and discussion. The CEO will sometimes invite relevant speakers and thought leaders to communicate changes in operations, lead discussions on challenging topics, and more. Employees are encouraged to join live, but the meeting is always recorded and available for employees to listen at their convenience. When the organization was building and implementing its AI governance framework, the CEO held several meetings to discuss AI-specific topics, such as the importance, value, risks, and challenges of responsible governance of AI, as well as any forthcoming operational changes.

> **"Leadership starts with the CEO and the board. The CEO has to exemplify 'say what you mean; do what you say' and set a consistent 'tone from the top' drumbeat on responsible AI and privacy by design."**
>
> **Caroline Louveaux**
> **Chief Privacy & Data Responsibility Officer, Mastercard**

1.2 Establishing organization-wide ethical principles – "The North Star"

Many organizations want to ensure and demonstrate that they are committed to leveraging the power of AI in a responsible way that would benefit all stakeholders equitably. To that end, they have established guiding principles to serve as a "corporate North Star" that fosters the responsible development and deployment of AI. These principles reflect the organization's perspectives, culture, and commitment toward AI ethics, responsibility, and accountability, and they often build upon and are aligned with the organizations' broader code of business or ethics. These principles serve as a compass that guides organizations when creating governance structures, implementing policies to develop and use AI responsibly, and dealing with escalations and difficult decisions and trade-offs.

There is an increasing convergence of these principles among organizations. Many had already established an initial version of guiding principles for organizational accountability, but they have since refined them (or are looking to refine them) to integrate specific principles regarding accountability in AI development and deployment. Common AI-related core principles include:

- Fairness and mitigation of risk of harms (to address bias and discrimination)
- Accountability and human oversight
- Safety, robustness, and reliability
- Transparency, explainability, and trust
- Privacy and security
- Integrity and respect
- Social good and benefit
- Sustainability

> **CASE STUDY 2: Including "Responsible AI" as a corporate and individual objective**
>
> A financial services organization included "Responsible AI," along with "Privacy by Design" and "Data Responsibility", as corporate objectives that every employee of the company must consider and uphold, regardless of function or role.

1.3 Establishing an AI and digital ethics oversight body

There is a growing trend for organizations to establish internal oversight bodies specifically for AI and digital ethics, such as review boards, advisory councils, or steering committees. The remit of these bodies is to support the ethical development and deployment of AI technologies within the organization and to provide relevant guidance as the final escalation point in decision-making, particularly in cases where the possible risks require heightened scrutiny. These bodies are becoming increasingly instrumental in managing high-risk technologies and data uses because organizations recognize that it is more crucial than ever to make ethics a core consideration at every point of the technology life cycle. Typically, their involvement in assessing the risk/benefit profiles of technologies and data uses is triggered by specific risk criteria set forth by the organization that will require a high-level review. These bodies are uniquely equipped to consider risks holistically from a broad range of perspectives and to weigh them against the benefits of the technology to the organization, individuals, and society.

The review bodies most often take the form of an internal, cross-functional council, but our study found that some organizations are exploring the possibility of an external advisory council or board.

Organizations uniformly stress the importance of bringing together expertise across a range of relevant disciplines (e.g., technical, social, industry, legal, ethics, HR, etc.). These individuals should be able to: 1. think critically about relevant ethical issues in the development, deployment, use, and sale of AI technology and associated data uses; 2. discuss and provide an opinion on matters that were referred to them for review; and 3. make informed and thoughtful decisions on how to proceed in a way that promotes responsible and accountable data practices. While there is no one "right" way to construct an advisory body, it is important to ensure representation from a diverse range of perspectives and expertise, with opportunities for meaningful participation in deliberations and decision-making.

> **CASE STUDY 3: Engineering as lead for the Responsible AI Governance Committee**
>
> A digital communications technology organization's senior leadership made a deliberate choice to designate the engineering team's representative to lead their Responsible AI Governance Committee. While the committee comprises representatives from several different relevant disciplines (e.g., IP rights, privacy, legal, engineering, etc.), the organization's leadership viewed engineering as the first line of defense and wanted to ensure that proper weight was given to their particular expertise.

> **CASE STUDY 4: Privacy as lead for the AI Ethics Board**
>
> The idea for an AI Ethics Board for an organization specializing in computer and information technology originated from their AI research team. However, the AI research team recognized that the organization's privacy team had the necessary expertise in creating comprehensive governance frameworks, identifying and mitigating risk, and working around existing and forthcoming regulation, and they intentionally invited the team to share its expertise and lead the initiative.

> **CASE STUDY 5: R&D and Enterprise as co-leads for the AI Resolutions Board**
>
> A pharmaceutical organization created an AI Resolutions Board to address issues regarding their accountable AI governance program. Senior members from the Research and Development team as well as the organization's AI enterprise team co-chair this board and have recruited representatives from relevant business functions (e.g., technical team, commercial, governance, risk, data, etc.).

Our research shows that most organizations are first choosing to establish an internal body, leaving it open as to whether in the future an external body might follow. They want to ensure that the internal body has sound, iterative processes and structures in place before considering external oversight. Organizations that are still in the early stages of AI governance feel the need to finalize and operationalize their internal policies and procedures before inviting external review.

For most organizations in the study, the decision to start with internal oversight bodies was taken because such bodies:

1. Comprise employees familiar with the organization's business, culture, and operations and are best positioned to identify, discuss, and address organization-specific ethical issues;
2. Can be more easily tailored to an organization's specific needs, as well as the industry it operates in;
3. Can still integrate additional perspectives by engaging external consultants to provide ad hoc expertise;
4. Give the organization greater control over the development and implementation of accountable AI practices throughout the organization because they include representatives from all relevant branches of an organization;
5. Provide space for frank and open discussions without fear of exposing intellectual property, trade secrets, or commercially sensitive information to those outside the organization.

Still, as mentioned above, organizations recognize the potential benefits of external advisory bodies that can provide independent oversight and feedback, and some are exploring ways to implement them in the near future. In any case, all organizations understand the value of AI and digital ethics oversight bodies, whether internal or external. Organizations believe that they can demonstrate accountability and commitment to responsible governance of AI and should be encouraged and incentivized further as a best practice.

> **CASE STUDY 6: Creating a cross-functional "AI Squad"**
>
> After the rapid public uptake of generative AI, a technology company established a cross-functional, internal group – known as the "AI Squad" – to tackle internal gaps, challenges, and concerns related to AI in a proactive manner. The group comprises representatives from AI research, product, compliance, security, legal, and privacy to ensure diverse input and collaboration.

> **CASE STUDY 7: Creating a region-specific external advisory board for adoption and awareness of "Responsible AI"**
>
> An organization that develops and provides AI technologies for businesses established an external AI advisory board specifically for one of its business regions. The advisory board gathered key advisors from academia, industry, and non-profit organizations to tackle the region's most pressing AI issues. The board meets regularly to work on various initiatives and challenges to integrate AI in an ethical and responsible way.

> **CASE STUDY 8: Creating external advisory boards for responsible AI governance**
>
> Some organizations in our study have established or are seeking to establish external advisory boards that will address responsible AI governance in its broader remit of organizational accountability, responsibility, and sustainability. Typically, these external bodies will include independent experts from industry, academia, civil society, and sometimes members of the general public who have applied to be a part of the oversight body. Members will often be given focused discussion topics or case studies and have opportunities to deliberate with senior-level executives.

Finally, organizations stressed that once set up, these oversight bodies must have the necessary authority to make impactful decisions, shape policy, build governance, and ultimately promote responsible AI practices within the organization.

1. Organizations that have successfully set up internal AI and digital ethics oversight bodies stated that some of the key success factors in establishing such councils were:
2. Defining the oversight body's scope;
3. Creating clear pathways for reporting, gathering, and escalating ethical issues to the oversight body;
4. Providing teams and employees within relevant business functions with the proper training to identify, consider, and escalate ethical issues; and
5. Giving the oversight body the necessary authority to make impactful decisions, shape policy, build governance, and ultimately promote responsible AI practices within the organization.

1.4 Creating a centralized governance structure

Nearly all organizations that participated in our mapping exercise are global enterprises with numerous business functions that develop and/or deploy AI in a variety of environments. Thus, many of them have created an overarching, centralized AI governance structure that provides strategic direction, support, oversight, and common standards that can be tailored in a risk-based way by business functions to what is appropriate, necessary, and most effective in their day-to-day operations.[10]

10 IBM's AI ethics governance framework serves as an effective example of how establishing an internal governance framework can help streamline the process of identifying and managing ethics concerns arising from AI projects. More details available at https://www.ibm.com/blog/a-look-into-ibms-ai-ethics-governance-framework/

It appears, therefore, that rather than relying on completely centralized or decentralized approaches to AI governance, most organizations tend to apply a more hybrid approach that includes high-level standards and centralized ethics oversight at the top, while allowing for flexibility in addressing AI issues within unique business verticals. From our study, this approach appears to be better suited to the evolving nature of AI technology and enables greater flexibility in the actual process of developing and deploying AI systems. Some organizations also borrow methodology from existing privacy and information security corporate compliance programs and implement a traditional "three lines of defense" approach.[11]

[CASE STUDY 9: Assigning responsibility over AI use cases to specific functions or teams] One organization allows individual teams greater independence by assigning "first line of defense" responsibility to the team or environment that is using a given AI application. For example, this organization's HR team uses an AI tool to filter resumes and aid in the recruitment and hiring process. Thus, the HR team is responsible for conducting the initial risk assessment to ensure that the AI application is safe and suitable for this use case. The team is also responsible for auditing the results and ensuring human oversight to see whether the resumes picked by the AI tool are those that would have been chosen by a human team.

1.5 Leveraging privacy programs and roles – expanding the role of privacy beyond legal compliance to include data ethics and digital trust

Our research revealed several key trends in how organizations are establishing their responsible AI governance programs vis-à-vis existing privacy and data protection programs and teams:

1. Given that privacy is a common thread connecting multiple data-rich business functions (e.g., security, HR, marketing, product, emerging technologies, business innovation, policy, etc.), **Chief Privacy Officers (CPOs) and their teams have become a natural first point of contact for various AI compliance and regulatory queries.** Many executives believe that data privacy teams are uniquely positioned to shoulder these new AI responsibilities and lead the charge in growing and maturing organizations' AI accountability programs. Their rationale is based on the following:
    - With ongoing AI regulatory developments and numerous others forecasted in the near future, data privacy professionals are already accustomed to implementing evolving laws that impact data and technology use and considering multiple, overlapping regulatory schemes.
    - Many accountable AI programs draw from practices familiar to data privacy management programs, such as data protection impact assessments and risk assessments. Moreover, the framework for implementing data privacy programs—such as the elements of accountability and many of the controls, tools, and processes—can be replicated in responsible AI programs. Data privacy professionals can thus leverage their experience to create consistent structures for accountable governance of AI.
    - **Instead of creating an entirely new AI governance framework from scratch, many organizations are assessing how their new accountable AI practices can leverage and build upon existing processes.** This also makes it easier for the workforce to adopt the new processes and lowers the barrier for implementation. Because many organizations already have a strong data privacy compliance function, teams responsible for that function can serve as a logical starting point for the development of an AI accountability program.
    - **Organizations are recognizing the need to eliminate internal silos of different digital and data compliance frameworks and apply a consistent, unified approach to data compliance.** A unified approach enables better reporting to the C-suite and Board and more effective internal and external messaging.

11 The three lines of defense (3LOD) model is a risk governance framework that defines roles and responsibilities for operational risk management into three distinct lines. Generally, the first line of defense owns and manages risks by managing internal risk controls and procedures on a daily basis; the second line of defense supports the first line of defense by helping to ensure and monitor effective risk management and compliance; the third line of defense evaluates the effectiveness of both the first and second line of defense by conducting internal audits. See more on the 3LOD model applied in AI governance in section 6.2.

- Many CPOs have already started to **expand their roles to include a broader remit of data and digital ethics, data regulation, and digital trust,** which many believe should naturally include AI.

2. In some organizations, when the CPO's role was expanded to include all matters regarding AI, the team responsible for the development of AI governance was also brought under the CPO's remit. However, the AI team was still able to function independently from the privacy teams.
3. In some organizations, responsible AI teams sit outside the privacy teams, but they still closely interact with them and leverage their expertise and existing data privacy tools and frameworks. However, as AI issues become more complex and intertwined with additional areas of law and policy, such as fundamental rights, intellectual property, cybersecurity, online safety, and content moderation, organizations are recognizing that privacy teams may not possess the expertise to address all of these topics on their own. In an effort to remain nimble and address the expansion of required skill sets and expertise, organizations are offering professional growth opportunities and encouraging teams either to upskill existing employees or hire additional members to further diversify the skillsets within their teams.
4. All organizations recognize the need for cross-functional, multidisciplinary, and multi-skilled teams that include data and AI scientists, researchers, and engineers. These members play an essential role in ensuring Responsible AI by Design, or the responsible development and deployment of AI technologies with privacy, safety, security, and transparency in mind. Their input in creating the overall AI governance structure is invaluable and should be included alongside the more traditional roles (e.g., ethics and compliance, legal, business) to build and deliver effective, responsible AI practices on the ground.

> **CASE STUDY 10: Enterprise driving the uptake of responsible AI practices**
>
> One organization shared that the impetus to develop and adopt responsible AI practices in a timely manner came from its commercial teams rather than its privacy and compliance functions. The "business" felt it could not wait for legal frameworks and AI regulations to become formalized because it was facing high expectations from customers to deploy AI-powered tools responsibly. Therefore, the enterprise took the lead in creating its internal processes for ensuring AI accountability and brought in the privacy and compliance functions afterwards to assist in its adoption.

2. Risk Assessment

CIPL has been a proponent of adopting a risk-based approach for managing and regulating the development of AI. Such an approach promotes protective, mitigating measures that are proportionate to the likelihood and severity of the risks of harm while enabling the benefits of AI technologies.[12]

As an integral aspect of organizational accountability, a risk-based approach to AI governance includes the following:

- **Building, implementing, and adapting AI programs and governance over time based on the relevant risks to individuals, society, and the organization itself.** Higher risks may require different correlating controls, processes, and tools. Changes in risk profile and external factors –such as new harms, new legal developments, and new concerns – may necessitate changes to the AI risk management program.
- **Systematically assessing both the risks and benefits** of individual products, projects, and deployments of AI technology and mitigating any identified risks.

As noted above, comprehensive risk assessments and analyses should consider:

- Risks to individuals, communities or groups of individuals, and broader society;
- Risks to organizations; and
- Benefits from the use of AI technologies for wider society, individuals, and organizations.

12 CIPL White Paper on 10 Recommendations for Global AI Regulation, October 4, 2023, available at https://www.informationpolicycentre.com//cipltenrecommendatioglobalairegulationoct23.pdf

To comprehensively assess AI risks, many organizations are taking a broadly centralized approach while still allowing some level of independence at the project level. This means they are providing ample support from the top, encouraging both horizontal and vertical lines of input and communication, and developing detailed risk assessment procedures to standardize the process.

Importantly, because the regulatory environment for AI is in flux while at the same time impacting several other areas of law (e.g., privacy, intellectual property, security, safety, anti-discrimination, etc.), there is a shared understanding that the frameworks for evaluating AI technologies and structuring AI governance must be continuously evaluated. Organizations should be nimble to stay abreast of the changing laws and embed new requirements as needed.

Additional practices for ensuring holistic, comprehensive risk assessments may include the following:

- Developing algorithmic impact assessments or fairness assessment tools to monitor and continuously test algorithms to avoid human bias, unfair discrimination, and "concept drift" throughout the entirety of the AI lifecycle;
- Requiring AI risk assessments at multiple points throughout the AI lifecycle, particularly for new or updated use cases or applications;
- Creating ethics, human rights, and/or data protection impact assessments;
- Creating a risk taxonomy that categorizes AI-related risks and allows for uniform assessment;
- Keeping a centralized repository of all risk assessment documentation;
- Developing standardized risk assessment methodologies that consider the benefits of the AI application or use, the likelihood and severity of risk factors on individuals and/or society, the level of human oversight needed for individually automated decisions with significant impact (e.g., legal ramifications), the ability to explain the technology in the appropriate context, and the ability to audit its effectiveness;
- Documenting considerations (e.g., accuracy, data minimization, security, transparency, scope of impact, benefits to society) for high-risk processing;
- Assessing data quality against key performance indicators (KPIs);
- Evaluating the data vis-à-vis the purpose of its use (i.e., the quality of the data, its provenance, whether it's personal, synthetic, in-house, or externally sourced);
- Developing frameworks for data preparation and model assessment – including feature engineering, cross-validation, back-testing, standardized KPIs;
- Enabling close collaboration between business and data experts (e.g., data analysts, data engineers, IT, and software engineers) on a regular basis to assess accuracy, ensure appropriate outputs, and allow for proper use of the model;
- Using privacy enhancing technologies (PETs) to preserve the privacy and security of AI systems; and
- Outlining escalation pathways to send AI-related issues to an AI ethics council or other oversight body.

2.1 Taking a centralized but flexible approach towards AI risk management

To manage AI risk in a consistent and repeatable manner, many organizations have implemented a centralized approach with added flexibility by allowing risk triage, assessment, and mitigation to be handled at the individual business functions or project level.

> **"A Responsible AI team (either centralized or federated) is a valuable resource for product teams and business units across the organization. It provides the standards and frameworks, as well as ongoing assistance and guidance."**
>
> **Addie Cooke**
> **Global AI Public Policy Lead, Google Cloud**

Factors used to develop a successful risk management framework include:

1. **Securing leadership support and engagement** Support from executive leadership and the C-suite is key for any organization's approach to AI risk management. Endorsement from the top signals that risk management and critical thinking regarding AI-related issues are expected across the organization. It also helps reduce confusion over who "owns" the risk management process, since it becomes a shared responsibility. Finally, it is essential that the senior leadership steers the organization's approach to risk, including the organization's overall risk profile and philosophy.
2. **Recognizing risk as a shared responsibility** Employees of an organization are crucial to the successful implementation of any AI risk assessment framework. Organizations stress the importance of any new AI risk assessments and processes being created with input and support from employees who will use them in their regular business functions. Relevant teams can work in tandem with the centralized governance team and contribute their expertise to create a framework that works within and alongside existing processes. While the central AI governance team can serve as a resource and facilitator of AI governance throughout the organization, ultimately, each business function or team working on AI development and deployment should feel empowered to use and adapt the framework for its own work. By engaging employees from relevant business functions early in the implementation process, organizations can ensure that the risk assessment framework is integrated into existing processes. This also encourages a sense of ownership and shared responsibility for the AI governance work and controls.
3. **Training relevant teams** Risk assessments cannot be effective if team members are untrained and unaware of how the new processes fit into the existing governance structure or how to think critically about AI-related risks. Therefore, organizations have tried to ensure that all relevant teams receive adequate training to assess and mitigate AI-related issues. Many organizations have also appointed business or project leaders to "own" certain risks of their AI application to contribute to the sense of responsibility and ensure that the framework is properly implemented throughout the AI lifecycle.
4. **Streamlining where possible** To avoid confusion that might result from different risk assessment criteria and documentation requirements for various business lines, organizations are aiming to establish centralized and simplified frameworks that combine AI governance, ethical principles, risk management, procurement, and other relevant processes. They are also looking to create an automated process that can help streamline AI risk assessments, while still recognizing the importance of human consideration and decision-making, especially for complex and high-risk AI applications. Finally, to drive consistency and document their processes, organizations are also establishing and maintaining a central repository for all records related to the AI risk assessment process.

CASE STUDY 11: Centralized repository of AI products and related risk assessments

Several organizations have created centralized inventories that not only store records of all of the organizations' existing AI projects, products, and services, but also log and track all risk assessments and send notifications to relevant teams when it is time for them to complete periodic risk assessments as mandated by company policies regarding AI.

CASE STUDY 12: Establishing a responsible AI program management team

A technology company established a team to monitor its massive AI product pipeline and ensure proper adherence to its governance framework across the entire organization. No AI product can be launched until it has gone through the risk assessment procedures and been reviewed by the program management team. Any mitigations needed for AI products must be done with the team's involvement and oversight.

2.2 Methodology and Key features of an AI Risk Assessment

Most organizations interviewed by CIPL are already in the process of or have already adopted specific processes to address AI risks. Organizations are taking a comprehensive, holistic approach to identifying and assessing AI-related risks and are designing and adopting processes to reflect this approach. To this end, **most organizations have adopted** a risk-based **approach that assesses the potential benefits and harms of AI technologies in the context of specific use cases, rather than assessing the benefits and harms of the technology itself,** and then triaging for severity and likelihood of any resulting harm.

Such assessments require an understanding of the definition of risk and the applicable risk-assessment obligations under the relevant regulations, as well as the consideration and implementation of effective mitigations and controls for the identified risks. Whether an initial risk "triage" is sufficient or whether a more robust risk assessment is required can depend on the complexity of the application and the intended use.

CIPL's research and interviews identified the following features and methodologies for the development of AI risk assessments in the context of AI:

1. **Creating a risk taxonomy**
    - Most organizations agree that the core consideration of any AI risk assessment should be the potential impact of the AI application on individuals and groups of individuals.
    - Organizations also recognize the importance of creating a taxonomy of AI-related risks to categorize them and allow for uniform assessment, and some have even begun the process of doing so.
    - Increasingly, organizations are taking a comprehensive, holistic approach to risk by considering the impact on individuals, groups of individuals, broader society, the environment, and the organization. Some have also begun to include AI issues within the remit of corporate ESG goals.[13]
    - Most organizations state that it is important for any taxonomy to mirror those set out in proposed regulatory frameworks or standards.[14] Yet, despite these efforts, there is still work to be done to create a broader consensus of actual AI-related risks and harms and how to uniformly assess the likelihood and severity of those risks by engaging with relevant stakeholders (including other organizations, civil society, regulators, and academia).
    - Organizations are still working on ways to define, track, and measure their progress in responsible governance of AI. In the absence of global standardization around key performance indicators (KPIs), organizations are performing regular internal audits or reviews. Externally, they are collaboratively participating in benchmarking

13 NIST AI Risk Management Framework, January 25, 2023, available at https://www.nist.gov/itl/ai-risk-management-framework; Putting principles into practice: How we approach responsible AI at Microsoft, Microsoft, 2020, available at https://www.microsoft.com/cms/api/am/binary/RE4pKH5

14 Organizations have most commonly referenced the EU AI Act, NIST AI Risk Management Framework, US Executive Order 13960, or ICO Guidance on AI and Data Protection.

studies and engaging in events (e.g., panels, workshops, roundtables) with relevant stakeholders to foster dialogue and encourage harmonization of standards and regulation on AI risk assessments.

2. Designing a comprehensive AI Impact or Risk Assessment

Internal AI impact or risk assessments are an essential and valuable tool for organizations to ensure that AI systems are developed and deployed responsibly. They enable organizations to systematically identify and characterize AI-related risks and mitigate them using proportionate measures.

The main objectives for an AI risk assessment that can effectively assess and mitigate AI-related risks are to:

- Identifying whether a particular application is considered AI or within the scope of the organization's definition for AI;
- Identifying the intended use(s) of the AI application and the "problem" the application intends to solve;
- Triaging to determine the relevant level of risk (e.g., high, limited, low);
- Determining the likelihood and severity of risk factors;
- Identifying and weighing the relevant risks against potential benefits from using a particular application; and Creating a plan of action to mitigate potential risks.

If an AI application requires further review before it can be approved for use, many organizations will refer the case to their AI and digital ethics oversight body for additional consideration and to ultimately either approve or deny the AI application.

3. Assessing the "likelihood and severity" of AI-related risks

Organizations carrying out systematic AI risk assessments seek to categorize risks according to the "likelihood and severity" of potential harms considering the available mitigations. This means that mitigations should be factored into the assessment of the risk. Thus, where the likelihood and/or the severity of the risk of harm is deemed to be high even after mitigations have been taken into account, the AI application should be scrutinized further internally, by referral to a higher-level advisory and oversight committee, such as an AI and digital ethics body (see 1.3).

The following likelihood-severity assessment matrix illustrates this approach[15]:

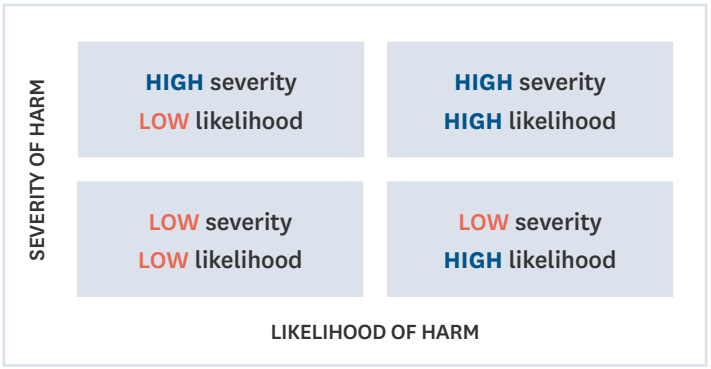

Figure 2. Risk Severity-Likelihood Matrix

Source: CIPL, adapted from Singapore Personal Data Protection Commission

15 Singapore PDPC's Model AI Governance Framework (2nd Ed) features a similar matrix on probability and severity of harm.

4. Implementing mitigation measures to address identified AI-related risks

Once potential risks have been identified, organizations can employ a variety of methods to mitigate them, such as:

- **Implementing specific additional safeguards** These include targeted controls, such as placing greater restrictions on high-risk use cases, tightening security, and requiring audit trails.
- **Keeping humans in the loop (HITL)** HITL requires incorporating human oversight, intervention, and decision-making throughout the AI lifecycle to ensure that AI systems are operating responsibly and in the intended manner. HITL can help build trust in AI technologies, particularly applications that operate in high-risk areas (e.g., financial services, insurance, employment, etc.).[16]
- **Implementing procedural and technical controls** Technical solutions, such as Privacy Enhancing Technologies (PETs) and Privacy Preserving Technologies (PPTs), can play an important role in maximizing the benefits of AI while preserving privacy and security by integrating them into the design and architecture of AI systems.[17] These technologies – including homomorphic encryption, federated learning, trusted execution environments, secure multi-party computation, differential privacy, and synthetic data – are gaining increased attention and investment from both industry and regulators as options for addressing privacy concerns in AI models.
- **Deferring to an AI and digital ethics body** Any AI use case deemed to be high-risk despite mitigations warrants additional review by an AI ethics council or a similar oversight body. Following review, the body can approve, discontinue, or suspend the development or deployment of that use case until further mitigations have been taken (see section 1.3 in Leadership and Oversight).

CASE STUDY 13: Using synthetic data to protect customers

Financial services deal with large amounts of personal data and acquiring such data directly from customers can pose great risks. One financial services organization has eliminated the need to collect data from customers by creating and using synthetic data to train their AI systems and test for fairness and bias.

CASE STUDY 14: Taking a product out of production until the risk has been properly mitigated

Where an organization was developing an AI-powered tool to create virtual backgrounds on a videoconferencing app, early research revealed that the feature was not performing well for certain hair textures and styles, especially under specific lighting conditions. The organization waited to release the feature until it was able to fully address the issue by retraining the AI model on more representative data sets to ensure a more representative, equitable, and inclusive user experience.

CASE STUDY 15: Maintaining a "Mitigation Library"

One technology company created an internal database known as the "Mitigation Library" to store documentation on previously employed mitigation measures for various AI-related risks.

16 There are emerging standards and legal requirements for keeping humans in the loop for AI systems, and different jurisdictions will enforce different obligations upon companies depending on how involved humans are. For example, the Colorado Privacy Act make the following distinctions regarding "profiling", or the automated processing of personal data: 1. "Human Involved Automated Processing" means the automated processing of Personal Data where a human (1) engages in a meaningful consideration of available data used in the Processing or any output of the Processing and (2) has the authority to change or influence the outcome of the Processing; and 2. "Human Reviewed Automated Processing" means the automated processing of Personal Data where a human reviews the automated processing, but the level of human engagement does not rise to the level required for Human Involved Automated Processing. Reviewing the output of the automated processing with no meaningful consideration does not rise to the level of Human Involved Automated Processing.

17 See more examples and use cases of PETs and PPTs in practice in CIPL's Privacy-Enhancing and Privacy-Preserving Technologies: Understanding the Role of PETs and PPTs in the Digital Age, December 12, 2023, https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl-understanding-pets-and-ppts-dec2023.pdf

CASE STUDY 16: An internal sandbox for high-risk AI experimentation

A pharmaceutical organization has created an internal sandbox to test and experiment with potentially high-risk or sensitive use cases of AI, including generative AI. In the sandbox, employees are able to use novel, generative AI tools to innovate and experiment not only with day-to-day business processes (e.g., creating slide decks, beta-testing new products, etc.), but also with research and development for new products (e.g., deducing new molecules for drug discovery) without having to undergo formal risk assessment procedures. However, employees are not allowed to use such technologies on any work that has an impact on or interacts with external stakeholders, such as patients (e.g., using a generative AI-powered chatbot to provide drug information to patients).

5. Implementing multiple "safety checkpoints" to ensure comprehensive risk assessments

Many organizations implement multiple "checkpoints" or continual assessments, as appropriate, throughout the entirety of the AI lifecycle to ensure ongoing identification of potential risks as they arise or become apparent. This includes:

- Mandating a formal risk assessment either at the earliest stages of the design process (if they are developing AI systems), or before any AI product can be approved for deployment. This ensures that any potential risks are caught early on and properly addressed before progressing too far where other competing issues, such as speed and cost, can come into play.
- Requiring new risk assessments to be done for any new use cases or applications of an existing AI product.
- Assigning the responsibility of performing risk assessments to a specific team or individual in order to maintain efficiency and consistency, while accommodating for multiple checkpoints.

CASE STUDY 17: Using an AI chatbot as the first risk assessment checkpoint

An organization programmed an AI chatbot that acts as an initial triage for assessing AI-specific risks for each use case. The self-assessment embedded into the chatbot takes employees who are developing or deploying AI through 11 questions based on relevant risk areas given the organization's AI principles. Upon completing the assessment, employees are directed to pursue further recommended actions to mitigate potential risks.

2.3 Leveraging existing expertise and processes for AI risk assessments

Organizations generally take a multistakeholder approach, drawing from diverse expertise and existing processes, when creating risk assessment frameworks for AI. As organizations design and adopt these frameworks, they are considering multiple policy areas and regulatory requirements simultaneously, notably privacy, human rights, cybersecurity, safety, intellectual property, and misinformation.

- Many organizations are leveraging their privacy teams' extensive experience with performing privacy impact assessments (PIAs) and complying with various sectoral regulations on top of privacy laws (e.g., finance, insurance, health, etc.)[18] and are appointing them to either lead or support building a similar risk-assessment process for AI-related risks.
- Some organizations are drawing from similar impact assessments (e.g., human rights, environmental, social, etc.) to further inform their internal AI risk management processes and ensure that they are taking a holistic approach in managing and mitigating AI risks.
- Some are trying to create a single risk assessment template to cover all relevant risk areas to streamline their internal risk assessment framework and to relieve business teams from the burden of performing multiple risk impact assessments for multiple regulatory areas.

18 CIPL Organizational Accountability Project, available at https://www.informationpolicycentre.com/organizational-accountability.html

- Many organizations recognize that it is now more important than ever to consult engineering or technical teams who understand and can communicate the capabilities and limitations of the AI technology as early in the process as possible to ensure that any risk assessment framework is useful and suitable from a technical perspective.
- Many organizations are also implementing an "ethics by design" approach (mirroring the "privacy by design" and "security by design" approach), which is a method of incorporating ethical principles to ensure that ethical issues are addressed early and reviewed throughout the development and design processes. [19] For example, organizations are implementing the ethical principle of fairness by creating fairness or ethics impact assessments to ensure that their AI systems do not create or perpetuate algorithmic bias, are universally accessible, and mitigate negative social impacts while offering their benefits equitably across social groups.

CASE STUDY 18: Testing accountable AI policies against the NIST AI Risk Management Framework

Meta and Accenture have launched a collaborative, policy prototyping program in the US by inviting a consortium of organizations to test the AI Risk Management Framework (AI RMF) developed by the National Institute of Standards and Technology (NIST). Specifically, they seek to test the efficacy of the framework in managing and mitigating risks associated with the development or deployment of generative AI.[20] The program will aim to inform the next iteration of the NIST AI RMF through rigorous, evidence-based testing on real, tangible issues facing organizations that are developing or deploying generative AI. It also seeks to enable interoperability and standardization of global AI frameworks.

CASE STUDY 19: Ensuring high quality data use in AI algorithm training

An organization that developed its own generative AI tool reduces the risks of unfairness, bias, and other harms by using high quality data that has been carefully and rigorously cleaned and curated through a proprietary process that includes a combination of thousands of separate automated and manual checks.

3. Policies and Procedures

Ideally, all elements of accountability implemented in an organization's AI program are reflected in its written policies and procedures. Thus, organizations are starting to establish internal, written AI policies, procedures, and tools that operationalize ethical principles, standards, and legal requirements into concrete actions, processes, and controls. These policies also build on an organization's internal rules, values, and priorities. Not only is written documentation helpful in encouraging responsible development and deployment of AI within the organization, but it also demonstrates the organization's commitment to accountable AI practices to external stakeholders, such as regulators, auditors, AI experts, business partners, investors, external oversight bodies, and end users.

Our research revealed that organizations have written policies and procedures in place that cover all aspects of accountability in AI programs. Below are a few examples of the scope of such policies:

- Adopting specific AI policies and procedures on how to develop, deploy, or sell AI;
- Drafting policies on the application of privacy and security by design principles throughout the AI lifecycle;
- Setting rules on the level of verification for data input and output;
- Requiring pilot testing of AI models before release;
- Specifying the use of protected data (e.g., encrypted, pseudonymized, tokenized, or synthetic data) in training AI models;
- Creating a glossary of AI-related terms for internal use and reference;

19 IBM elaborates more on the importance and methodology of an "Ethics by Design" approach, and provides a case study of the approach used in practice: https://www.ibm.com/cas/7PWAWRQN

20 Meta Open Loop's first policy prototyping program in the United States, available at https://www.usprogram.openloop.org/

- Promoting the use of smaller, higher quality data sets;
- Cleaning and curating data sets before model training through automated or manual checks;
- Considering relevant and appropriate use of PETs and PPTs to integrate privacy and security controls into AI models;
- Outlining special considerations for organizations creating and selling AI models, software, applications;
- Developing a fairness or AI impact assessment to analyze and mitigate AI-related risks;
- Creating due diligence/self-assessment checklists or tools for business partners deploying AI;
- Clearly defining escalation steps for reporting high-risk AI issues;
- Implementing an ideation phase with all stakeholders (e.g., data scientists, business, final user, control functions) where needs (including explainability), outcomes, validations rules, maintenance, and budget are discussed;
- Implementing specific policies for internal GenAI use;
- Requiring consideration for diversity in relevant teams and business functions;
- Implementing internal policies in parallel with forthcoming AI regulation;
- Translating internal principles-based policies to third-party vendor agreements, language, and due diligence processes; and
- Creating processes for review of high-risk AI use cases by an AI ethics board or council.

3.1 Agreeing upon a common language and scope

While there are many ways for organizations to define AI and related terminology, depending on various business functions, needs, products, or services, organizations are fostering consistency and coherence by creating taxonomies for internal reference and use, and by looking to forthcoming regulation.

Almost all organizations in our research stated that it has been a challenge to agree on a common taxonomy of AI-related terms to use across business functions or between organizations. Not only are there regulatory and policy debates on how best to define and regulate AI, but organizations report that what falls under the scope of AI internally can change depending on role, context, and level of expertise. For example, technical teams (e.g., engineering, data science) may use a more specific and granular definition of AI than wider enterprise teams (e.g., marketing, policy). Furthermore, different sectors may place different emphasis on certain characteristics or applications of AI than others, which in turn informs the scope of their definition of AI. For example, an insurance organization may limit its definition of AI to existing regulations that cover automated decision making (ADM) and algorithmic fairness and bias, whereas an organization that builds its own AI tools may require a more specific, technical definition of AI. Furthermore, while AI applications are not new, the leap in generative AI and the resulting democratization of GenAI tools have created an additional need for organizations to determine which AI technologies fall within the scope of their responsible AI programs and therefore, are subject to their new controls, policies, and procedures.

Nevertheless, many organizations have already established their own definitions of AI in their internal policies or have created a glossary of technical terms for internal reference. Most commonly, they have referred to definitions put forth by credible organizations, such as NIST and the OECD, in anticipation of potential convergence to standards or regulations in the near term. There is also widespread agreement among organizations that they will keep a close watch on upcoming regulations, such as the EU AI Act, that could inform their internal definitions of AI. Finally, several organizations find it helpful to use the definition of automated decision making (ADM) under relevant laws, such as EU GDPR and US state privacy laws, to classify AI systems.

CASE STUDY 20: Maintaining a working glossary

Many organizations have created a working glossary of AI and data-related terminology to serve as a reference for their employees. They recognize that establishing a bedrock of standard language is imperative to encourage collaboration and understanding between different business functions within the organization. One organization, in particular, has published an online public glossary of terms commonly used in AI concepts and technologies.

3.2 Putting company-wide principles into practice

As discussed above, many organizations have already established guiding principles for ethical development and use of AI that are consistent with their broader business ethics principles. Since principles are only as good as how they are operationalized, all organizations in the study are now working to build out their high-level principles into concrete, practical guidelines and requirements that can be easily adopted, are scalable across the organization, and are sustainable in practice. By tying ethical principles to specific guidance that can be integrated into various internal functions, organizations are seeking to ensure that responsible, accountable governance of AI is a shared responsibility. Additionally, by making the requirements principles-based rather than prescriptive, organizations can ensure that their policies and procedures are flexible enough to accommodate changes in technology and new regulatory requirements.

3.3 Creating policies in parallel with forthcoming regulation

Because the regulatory landscape remains in a nascent stage and AI technologies are being developed and deployed at a rapid pace, most organizations expressed that they cannot wait for compliance requirements to become formalized to create governance structures for AI. Thus, many have already created an "AI Policy 1.0" as a living document that operationalizes the company's responsible AI principles in a manner consistent with existing regulatory requirements, with the expectation that it will be updated as additional regulatory obligations are established.

3.4 Including a diversity of voices within the AI governance process

Organizations are recognizing the importance of bringing together multi-disciplinary and diverse perspectives to ensure non-biased, non-discriminatory, and responsible development and deployment of AI technologies and uphold the organization's brand and reputation.

Given the need to consider risks of harm, trade-offs, and ethical questions, it is crucial for teams working on AI development or deployment and AI governance to represent a wide range of voices in terms of fundamental diversity indicators (e.g., gender, age, religion, sexual orientation, race, and ethnicity), expertise (e.g., technology, social, industry, legal, human resources, government relations, ethics), and geography.

"It is important to ensure the diversity of teams that work on responsible AI matters – think of culture, age, and geography but also professional skills. That broader perspective will produce better results."

Florian Thoma
Senior Director, Privacy, Data & AI Compliance, Accenture]

3.5 Managing third parties and setting the standard for accountable AI practices

Since many organizations either procure AI products from or sell them to third parties, organizations recognize the importance of implementing accountable AI practices into business procurement and third-party management processes. Organizations demonstrate accountability by ensuring that their third-party vendors are held to the same standards and principles that they adhere to themselves, and that both parties are aligned in their commitment to accountable AI practices. This can be done through a variety of methods, such as:

- Including legal, procurement, and vendor management teams in policy-making discussions and processes;
- Educating relevant teams on broader accountable AI goals and how to operationalize AI governance guidelines and processes;
- Embedding AI review processes in existing third-party risk management or procurement processes;
- Updating third party contractual templates to reflect, encourage, or require compliance with an organization's internal principles or policies;
- Requiring transparent explanatory documentation on AI systems from AI system developers and/or deployers (e.g., model and system cards, technical reports) and from providers of AI products (see more in section 4.1);
- Requiring due diligence on responsible AI aspects before entering into any business agreement; and/or
- Including vendor management or procurement as an aspect of internal AI auditing practices.

Ultimately, management of third-party risk is still a developing area of accountable governance of AI and continues to be a focus of emerging AI regulatory requirements.[21] Many organizations agree that vendors are at varying levels of readiness to comply with such standards, particularly regarding transparency. Thus, it will be important for leading organizations to set an example for ethical and responsible practices for vendor management and procurement by reflecting their own internal principles in business decisions. Furthermore, organizations have expressed that as vendor certifications become more widely accepted and recognized, they will use certifications or codes of conduct as key due diligence tools to identify vendors that have demonstrated their commitment to accountable AI practices.

> **CASE STUDY 21: Updating third-party vendor or business-to-business contract policies**
>
> Many organizations have gone through a comprehensive overhaul of their third-party contracts and processes and updated them to reflect the organization's Responsible AI principles. They wanted to create internal obligations and standards for any vendor that wants to do business with them and encourage a "leveling up" of AI risk management and transparency for all players.

4. Transparency

Providing transparency is an essential aspect of building trust and garnering credibility in developing and deploying AI technologies, particularly with the widespread availability and use of GenAI. Transparency is also a key feature of emerging regulatory guidance and requirements. Truly accountable organizations strive to be transparent about their AI practices with both internal and external stakeholders.

Our research shows that organizations are thoughtful about how they deliver transparency measures and to whom they are directed. They are often tailoring the level and type of information provided based on the target audience. Below are some practices that organizations have implemented to deliver transparency to stakeholders:

- Tailoring transparency measures for the different needs of end users, regulators, business partners, and internal stakeholders at all stages of the AI lifecycle;
- Communicating disclosures in a simple, easy-to-understand manner;
- Considering how AI disclosures can be inclusive and accessible for those with special needs/disabilities;
- Establishing a transparency trail to explain automated decision-making and broad workings of algorithms;
- Providing notice when the system relies on AI/ML);
- Providing counterfactual information (e.g., how different inputs can affect the output of an AI model);
- Understanding customers' expectations and deploying AI technologies based on their readiness to embrace AI;

21 Section 10.1(xiii)(E) of the Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, October 30, 2023, available at https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/

- Implementing tiered transparency;
- Defining criteria for internal deployment of AI technologies based on usage scenarios and communicating them to users;
- Publishing model or system cards (i.e., short documents accompanying AI models that describe the context in which a given model is intended to be used and how the model performs in a variety of conditions);
- Creating a data hub for information regarding data governance, data accessibility, data lineage, data modification, data quality, etc.;
- Tailoring transparency to the identified risk (e.g., using watermarking for generative AI output) where possible and appropriate;
- Participating in benchmarking opportunities, public engagement, and regulatory sandboxes; and
- Using visualization tools to depict difficult, technically complex concepts to end users.

4.1 Taking a layered approach towards transparency

Organizations are taking a multi-pronged approach towards transparency by using foundational processes from privacy (e.g., consent notices, privacy statements, just-in-time notices, etc.) as well as more creative methods, such as FAQs, blogs, videos, webinars, and product-specific documentation that are tailored for specific audiences (e.g., model and system cards). The method of transparency depends on the role of the speaker (the one providing information, e.g., developer of the AI model) and the role of the recipient (the one receiving information, e.g., end users, organizational deployers, business customers).

For example, developers of AI technologies often publish transparent explanatory documents (e.g., model cards, system cards, technical reports) that provide essential information regarding an AI model, such as how it was built, how it works, a summary of the data it was trained on, its intended use cases and contexts, key limitations, and basic performance metrics.[22] Developers design these resources to make AI more explainable and transparent to external stakeholders thus building greater trust in their products. Developers also offer these documents with varying levels of technical information to cater to different audiences and levels of expertise. For example, developers of commercially available AI models may offer technically detailed model cards for business customers and their internal engineering teams, but organizations deploying AI technologies for use by the general public may publish a model card that uses plain language and helpful infographics to simplify technically sophisticated concepts.

Furthermore, organizations recognize that true accountability in AI transcends compliance; transparency allows organizations to build trust in their products and their business. In general, organizations target transparency tools and materials towards three different audiences: consumers/end users, third parties, and regulators.

1. Transparency to consumers and end users

Organizations recognize that if the AI technology is not accepted by the end user, they cannot deliver its full potential and benefits. Therefore, transparency to end users is a smart and responsible business decision to engender trust not only in the technology itself, but also in the organization. Organizations deploying AI typically provide the following information to end users:

- When AI technologies are being used;
- The capabilities and limitations of a given model;
- The data on which the model was trained;
- The data used to generate outputs;

22 Model Cards for Model Reporting, Mitchell, Wu, et al., October 5, 2018, available at https://doi.org/10.48550/arXiv.1810.03993

- Whether data is retained (and if so, what and for how long);
- Avenues to remediate or appeal outputs produced by the model; and
- Whether user choices can influence system performance.

Moreover, under relevant data protection laws, organizations may be required to provide information to the extent the model uses personal information and may fall within the scope of such laws.

CASE STUDY 22: Deploying the "WAIST" ad tool

An organization that uses AI algorithms to power "ranking and recommendation" features (i.e., advertisements) on its web platform deploys a "Why am I seeing this?" tool that provides users with the ability to view a brief explanation of the factors considered in displaying a specific advertisement and how the user may exercise control over the types of advertisement shown. Through this tool, users can provide feedback on whether the advertisement is relevant or appropriate and adjust relevant settings.

CASE STUDY 23: Incorporating transparency into a chatbot

A technology organization has embedded transparency disclosures and FAQ documents into its generative AI chatbot so that users can directly ask the chatbot how the service works, what kind of functionality they should expect (e.g., data processing, available controls), how it moderates content, and more.

CASE STUDY 24: Publishing an annual "progress report" of the implementation of AI Principles

Every year, Google publishes a progress report that details how it operationalizes its AI Principles to demonstrate transparency and show how it is holding itself accountable to its own AI Principles. Google's 2023 report discusses the application of its AI Principles to the research and development of its generative AI models.[23]

2. Transparency to business partners and B2B customers

Transparency to business customers is key, particularly for large AI developers. These organizations deploy a wide range of approaches to deliver transparency, including the model cards mentioned above, and are increasingly seeking opportunities to demonstrate greater transparency to the public, media, and AI experts. One that is growing in popularity is participating in public "red teaming" challenges (see more in section 6.3). Red teaming is the method of deploying ethical hackers to leverage the tactics, techniques, and procedures (TTPs) of cyberattackers to proactively identify security weaknesses and gaps.[24]

The most recent and notable public red teaming challenge took place in August 2023 at "DEF CON," one of the world's biggest annual hacker conferences. Nine AI labs leading generative AI development[25] opened their large language models to be publicly "stress-tested" by hundreds of participants from a variety of backgrounds, including students, white-hat hackers, academics, industry workers, and more. The results from this exercise are still pending but they are highly anticipated. Red teaming is increasingly expected to become a regulatory requirement and an essential element of organizations' accountable AI programs and monitoring and verification efforts.

23 Google, AI Principles Progress Update 2023, available at https://www.ai.google/static/documents/ai-principles-2023-progress-update.pdf

24 President Biden's recent Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence defines the practice of red teaming as "a structured testing effort to find flaws and vulnerabilities in an AI system, often in a controlled environment and in collaboration with developers of AI. Artificial Intelligence red-teaming is most often performed by dedicated "red teams" that adopt adversarial methods to identify flaws and vulnerabilities, such as harmful or discriminatory outputs from an AI system, unforeseen or undesirable system behaviors, limitations, or potential risks associated with the misuse of the system." The Executive Order also places requirements to report results of red teaming exercises for certain dual-use foundation models to the US government.

25 Participants included OpenAI, Anthropic, Meta, Google, Hugging Face, Nvidia, Stability.ai, Cohere, and Microsoft, and the event was supported by the White House Office of Science, Technology, and Policy, the National Science Foundation's Computer and Information Science and Engineering (CISE) Directorate, and the Congressional AI Caucus

For other organizations who sell AI products for third-party use, transparency to such customers is still a developing area (as stated in section 3.3). They are navigating how to build trust with business partners while upholding their own standard of commitment to accountability and ethics. Many have found success by updating contract clauses to reflect their own corporate AI principles and initiating conversations with vendors on what responsible AI partnership can and should look like.

3. Transparency to regulators

Organizations recognize that regulators expect an enhanced level of transparency. Regulators are increasingly looking for detailed, technical documentation on how organizations are complying with relevant regulatory and legal requirements in their AI development and deployment.

For most organizations, transparency to regulators has been mostly reactive and in response to on-demand regulatory requests. However, many are trying to create avenues for proactive engagement, such as the following:

- **Keeping detailed documentation** Organizations are applying a key lesson learned from data privacy to AI transparency by creating and maintaining detailed evidence and documentation of their AI policies, procedures, risk assessments, mitigations, and decisions. Not only is this a good business practice, but it is also anticipatory of potential required audits in forthcoming regulations.
- **Sharing information about products and offering opportunities for discussion** Organizations are increasingly trying to create opportunities for engagement with regulators by: 1. sharing information regarding some of their AI applications and products; 2. providing practical knowledge from relevant technical teams (e.g., AI technologists and engineers); or 3. sponsoring collaborative events with other industry players.
- **Submitting responses to public consultations or engaging in regulatory working groups** Organizations see public consultations and regulatory working groups as crucial opportunities to share their perspectives and opinions on forthcoming initiatives with regulators.
- **Participating in regulatory sandboxes** Regulatory sandboxes provide organizations with a test bed to apply AI laws to innovative products and services under the supervision of and with feedback from a regulator.[26] They also offer organizations the opportunity to provide feedback on the practicality of regulatory requirements, demonstrate how AI technologies work on the ground, and improve regulators' understanding of AI technologies. Nonetheless, organizations understand that regulatory sandboxes are for special cases and cannot be used as a substitute for fundamental transparency practices and engagement with regulators.

CASE STUDY 25: Tracing data lineage and provenance

An organization specializing in computer and information technology has created documentation processes to disclose information about the data used to train its AI models, such as the type of data used, where the data came from, the quality of the data, and how risks have been addressed or mitigated.

CASE STUDY 26: Using visualization tools

A financial services organization uses "visualization tools" to depict how their AI systems work without revealing sensitive information to bad actors that could potentially compromise the safety and security of their fraud models.

26 CIPL Paper "Regulatory Sandboxes in Data Protection – Constructive Engagement and Innovative Regulation in Practice, March 8, 2019, https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_white_paper_on_regulatory_sandboxes_in_data_protectionconstructive_engagement_and_innovative_regulation_in_practice__8_march_2019_.pdf.

CASE STUDY 27: Creating a UX framework for AI explainability

A technology company has developed a user experience (UX) framework that provides the appropriate level of transparency and explainability to a given audience (e.g., platform user, AI researcher, regulator, etc.).

5. Training and Awareness

Employee training ensures awareness of an organization's AI policies and procedures, embedding them into the culture of the organization over time. Mandatory trainings on AI ethics and the organization's AI program allow employees to understand how AI ethics principles translate into their daily roles and responsibilities, how they relate to regulatory requirements, and how they can cultivate a shared sense of responsibility and commitment to principles articulated by leadership. Specific practices adopted by the organizations in our study include:

- Providing specific training for data scientists and engineers, including how to address relevant ethical issues (e.g., how to limit and address bias);
- Creating opportunities for cross-functional training (e.g., between privacy professionals and AI engineers);
- Tailoring trainings regarding ethics and fairness in AI for relevant teams;
- Compiling and making available AI use case information where relevant risks have been mitigated or deployment has been halted;
- Creating a "translator" role that helps explain the impact and technical capacities and limitations of AI;
- Sharing case studies to help employees learn how to address potentially complex, ethically challenging AI cases; and
- Incentivizing compliance with completing ethics training by pairing it with eligibility for bonuses, pay raises, and/or promotions, or incorporating it into other mandatory training activities.

5.1 Providing internal AI ethics and governance training

Nearly all the organizations in our study recognize the need to operationalize their core principles and improve workforce awareness and compliance to their accountable AI program through AI ethics training. While many already require basic ethics and compliance training, some have created additional training modules specifically for AI ethics. These bespoke courses have been created either by an internal team of diverse experts or by an external vendor from industry or academia. All organizations stated that any effort to develop and deploy AI responsibly should include instruction on how to critically assess and address AI-related ethical challenges.

Additionally, many organizations are creating specialized modules for particular roles or business functions that have greater interaction and influence on AI development and deployment (e.g., ethics training through a technical lens for technologists and engineers; AI ethics and technical foundations courses for the C-suite). Creating tailored trainings for different audiences can help address the gap in data science and AI literacy that many technologists and data scientists report has become a significant barrier towards productive internal engagement regarding responsible AI practices.

While our research found that, currently, AI-related trainings are predominantly offered through online courses, many organizations are in the process of incorporating more active learning strategies, such as interactive workshops using real-world case studies, company discussion forums, panel discussions, and more.

CASE STUDY 28: Offering "Responsible AI" trainings

With "Responsible AI" as part of its broader corporate mandate, one technology company requires "Responsible AI" training for its entire workforce within its general business conduct training. The organization also offers role-based training to those who are engaged in AI-related work and bespoke trainings for GenAI and large language models.

> **CASE STUDY 29: Creating a central "learning hub"**
>
> A software company has a dedicated "learning hub" that every employee can access to get a solid understanding of AI and its role in various products offered by the organization. The hub houses training materials, such as videos categorized by topic, employee role, and competence level, as well as white papers and recordings of cross-functional training events.

5.2 Encouraging external AI ethics and governance training

Organizations have taken great interest in the propagation of certification programs designed to deepen employees' knowledge of responsible AI, AI ethics, and data ethics. Many universities and organizations have gathered teams of interdisciplinary experts to offer courses, programs, and workshops on these topics either in-person or virtually.[27] Professional organizations, such as the Institute of Electrical and Electronics Engineers (IEEE) and the International Association of Privacy Professionals (IAPP) have developed similar offerings (e.g., IAPP AI Governance Training Program, IEEE "CertifAIEd" Accreditation). These programs provide a foundational understanding of AI and its associated risks, the AI lifecycle, significant AI governance frameworks, and an overview of current and emerging laws applicable to AI systems, but some organizations are waiting for these certification programs to mature and become more broadly adopted before encouraging or requiring employees to obtain them. In the meantime, organizations have been referring to curricula from such programs to help shape internal trainings, since they recognize that the programs pull together values, principles, and knowledge critical to achieving accountability in AI development and deployment.

5.3 Fostering a corporate environment that encourages ethics as a shared responsibility

Many organizations are "upskilling" their current workforce, since "responsible AI" is still a relatively new and rapidly evolving field. Coaching employees to be intentional, thoughtful, and ethical is an important business investment. Some organizations are creating networks of AI "champions" or "ambassadors" to promote accountable AI practices as a part of a company-wide, shared initiative; these ambassadors are strategically placed in various business and product functions.

> **CASE STUDY 30: Establishing a network of "AI Ambassadors"**
>
> A software company has created an "AI ambassador network" where employees who are passionate about learning and sharing knowledge regarding AI and machine learning can earn certificates and badges for various activities that raise AI awareness within their respective business units (e.g., sharing AI-related news on their team channel, answering AI-related questions in their company chat, posting an AI-related blog or article).

As mentioned in section 1.1, it is critical for executive leadership to set a "tone from the top" by demonstrating a high level of commitment to AI accountability and aligning trainings with the organization's overarching ethical principles. Such actions encourage the workforce to view accountability in AI practices as a company-wide, shared responsibility.

> **CASE STUDY 31: Setting soft pressure to discourage noncompliance with ethics training**
>
> Executive leadership of one organization imposed punitive measures for noncompliance with mandatory AI ethics and compliance training (e.g., impact on bonuses, pay raises, and promotions) to demonstrate their seriousness and commitment to creating a workforce that upholds company values.

Additionally, many organizations are requiring mandatory training for their newly implemented AI governance frameworks and policies (e.g., providing an overview of AI ethics principles or explaining a new internal policy on GenAI) to keep everyone abreast and engaged.

Finally, many organizations report the benefit of conducting cross-disciplinary team training, where, for example, AI technologists can learn about relevant policy and legal considerations for a successful accountable AI program, and legal, compliance, and policy teams can learn the basic foundations of how AI technology works.

27 Some examples include the MIT Certification Program in ML & AI, Berkeley Law AI Institute, Stanford AI Professional Program, University of Oxford online courses, NVIDIA Deep Learning Institute, Google Cloud Training, and IBM AI & Machine Learning Training.

CASE STUDY 32: Creating AI project opportunities for non-AI teams

A financial services organization has designed opportunities for non-technical employees to participate in projects with AI-related components. These opportunities allow employees to familiarize themselves with internal processes and understand that responsible and accountable AI practices are a shared responsibility across the entire organization

6. Monitoring and Verification

Monitoring and verification of the implementation, internal compliance, and effectiveness of an organization's accountable AI program are essential to ensure that the accountability loop is closed. They enable organizations to demonstrate accountability both internally and externally, thereby earning trust from all relevant stakeholders. Many organizations view their accountable AI programs as nascent and are continuously working to learn which practices are most effective. Many have chosen to conduct internal audits and periodic reviews of their internal policies to test and respond to areas of potential noncompliance, noting that external audits or certifications may one day become required. Other practices adopted by the organizations in our study include:

- Incorporating "human in the loop" (HITL) in design, oversight, and redress;
- Identifying and understanding which business functions are using AI;
- Providing the capability for human audit of input and output;
- Ensuring human review of individual decisions with legal or similarly significant effects;
- Monitoring the data ecosystem—from data flow in, through data process, to data flow out;
- Using different auditing techniques;
- Deploying counterfactual testing techniques;
- Pre-defining AI audit controls;
- Creating an internal audit team with expertise in AI and other emerging technologies;
- Allowing human control or intervention where technically possible and reasonably necessary;
- Monitoring AI models (e.g., back-testing and feedback loop) and conducting ongoing maintenance; and
- Red teaming and adversarial testing of AI models.

6.1 Shared understanding of the evolving nature of the policies

While many organizations' AI governance programs are in an early stage of development and implementation, organizations are putting in place monitoring mechanisms to ensure that their programs remain effective and reflect the ongoing changes to the regulatory landscape. In this initial phase, many organizations are putting the focus on developing their internal policies and encouraging adoption throughout the organization. Most organizations believe that getting teams into the habit of incorporating these policies into their current roles and responsibilities will make refining the policies and identifying potential gaps in the policies easier in the long run.

Currently there is little global standardization or agreement on the elements of accountable AI programs or on the KPIs to measure and track the efficacy of such programs. Thus, as AI accountability programs mature, organizations recognize that to be a true thought leader in this space, they should not only demonstrate their successful adoption and advancement of internal AI practices, but also advance the global discussion concerning regulatory efforts and industry standards.

6.2 Conducting periodic internal audits and reviews of internal policies

Nearly all organizations in our study are using their internal audit function to conduct periodic internal audits and reviews of their accountable AI practices. Many organizations have found it helpful to adhere to the "three lines of defense" model, described below:

- First line of defense – Employees and teams working directly with AI products or applications act as the "first line of defense" by managing day-to-day operations specifically required by the accountable AI governance program (e.g., performing risk assessments, auditing the performance of AI products, etc.).
- Second line of defense – Internal AI governance teams act as the "second line of defense" by monitoring adherence to policies, tracking risk assessments, and reporting to the senior leadership regarding implementation efforts.
- Third line of defense – The internal audit team acts as the "third line of defense" by evaluating the effectiveness of the first and second lines of defense. This includes periodic reviews of risk assessment processes and the creation of unique auditing procedures specifically for AI. Organizations are aiming to set up a specific AI compliance or assurance function when processes become more formalized.

> **CASE STUDY 33: Establishing "Responsible AI Champions" throughout the organization**
>
> A telecommunications organization has appointed volunteer "Responsible AI Champions" who act as designated points of contact within a given business unit to review whether the appropriate AI governance procedures have been fulfilled, whether risk assessments have been performed comprehensively, and whether any issues should be referred to the company's AI advisory body.

In addition, some organizations task specific individuals with tracking global regulations to review proposed requirements and implement measures to address them pre-emptively. For example, many organizations have already begun to assemble written documentation of their programs in anticipation of regulations requiring compulsory external audits.

To the extent an internal audit finds compliance gaps or shortcomings, the AI governance team will create and implement a remediation action plan and report on its progress to all relevant teams.

6.3 Conducting internal and external red teaming

For many large organizations, particularly those that develop generative AI or foundational models or other large AI technologies and products, red teaming and adversarial testing of AI models is emerging not only as an expectation, as per the US Executive Order[28] (referenced in section 4.1), but also as a useful tool for verifying the effectiveness of internal AI policies and accountability measures and mitigating any risks. Some organizations have established internal "red teams," while others have outsourced red teaming to external networks or consultants. If a red teaming exercise reveals vulnerabilities, organizations will ensure that these findings are fed into a reinforcement loop where the company can take action to address the vulnerabilities and fine-tune systems before the technology or product is deployed to users. In addition to red teaming, organizations use other accountability measures for AI, such as creating a code of conduct for AI auditors, using training curricula, and promoting certification programs.[29]

> **CASE STUDY 34: Participating in public red teaming challenges**
>
> In August 2023, eight organizations leading generative AI development participated in DEF CON's Generative Red Team Challenge, where they opened their large language models to be publicly "stress-tested" by hundreds of challenge participants. [30]

6.4 Looking towards developing individual and organizational certifications

Organizations are very interested in the development and standardization of AI governance certifications issued by external or independent certifiers, or by means of a self-assessment against a set of standards or requirements.

These types of certifications are still in nascent stages and have yet to be widely mandated or recognized by regulators.

28 88 FR 75191.

29 The newly formed International Association of Algorithmic Auditors (IAAA) has been launched by academic and industry leaders in AI to establish guardrails, create consistency, and produce standards, methodologies, and tools for auditing AI.

30 Participants included Anthropic, Cohere, Google, Hugging Face, Microsoft, Meta, NVIDIA, OpenAI, and Stability AI, and the event was supported by the White House Office of Science, Technology, and Policy, the National Science Foundation's Computer and Information Science and Engineering (CISE) Directorate, and the Congressional AI Caucus.

However, many organizations foresee that both types of certifications will gain popularity once established. They believe that having such certifications could eventually drive business decisions by acting as a "seal" that accredits them as ethical, responsible, and accountable. Certifications could also streamline due diligence obligations in B2B relationships, with respect to certified vendors and other third parties.

> **CASE STUDY 35: BBB National Program publishes self-regulatory Principles and Protocols for Trustworthy AI in Recruiting and Hiring**
>
> In June 2023, BBB National Program's Center for Industry Self-Regulation's (CISR) policy incubator published "Principles and Protocols for Trustworthy AI in Recruiting and Hiring." [31] This self-regulatory program offers a global baseline standard for industry transparency when companies utilize AI tools in hiring processes. The Principles offer practical guidance for employers and application vendors for circumstances in which advanced, adaptive algorithms are used as part of the employment selection process, while the Protocols establish a baseline framework for independent certification to the Principles.

7. Response and Enforcement

The final element of CIPL's Accountability Framework, response and enforcement, addresses how an organization would:

1. Enforce its accountable AI program in cases of internal noncompliance;
2. Deal with security incidents and data breaches;
3. Respond to requests and complaints from individuals and end users; and
4. Respond to requests and investigations from regulators or external auditors.

Though most organizations have not yet had to respond to regulator inquiries or investigations regarding their accountable AI practices, many stated that they had already deployed multiple existing channels for both employees and customers to deliver feedback and ensure that findings from internal audits were comprehensively addressed. Some additional best practices include:

- Enabling redress mechanisms to remedy an AI decision;
- Permitting redress through a human, not to a bot; and
- Developing communication channels for internal (e.g., for employees) and external (e.g., end users, business customers) to report and address feedback, complaints, requests, etc.

7.1 Acting upon findings of audits and reviews

Once an internal audit of the AI program has been completed, organizations create a remediation plan that outlines actionable corrective measures to mitigate where the audit has shown gaps or deficiencies and assign roles and responsibilities to carry them out.

Some organizations have found it helpful to have a technician (e.g., software engineer, data scientist, etc.) assist in creating and enforcing corrective mechanisms, especially those that require specialized, technical expertise.

After the remediation plan has addressed and corrected the issues raised in the audit, executive leadership can be apprised of the completed actions.

31 BBB National Program's Center for Industry Self-Regulation, Principles for Trustworthy AI in Recruiting and Hiring, June 8, 2023, available at https://bbbprograms.org/media-center/bd/insights/2023/08/25/ai-principles-and-protocols.

[CASE STUDY 36: Establishing a bug bounty program

One technology company has created a "bug bounty" program to incentivize external researchers to identify and report potential security vulnerabilities for a monetary award. The program receives hundreds of reports per year, greatly improving the security and privacy protections of the organization's products. A GitHub repository collects, analyzes, and prioritizes the reported bugs so that the organization can address them in a timely and efficient manner.

7.2 Dealing with data or security breaches

The organizations included in our study have specific processes in place to manage security incidents involving their AI systems, including breaches of personal data. While many already have cross-functional response teams to handle privacy and security-related incidents, they are leveraging those same teams for any AI-related incidents. These teams, as well as any relevant employees (e.g., teams working directly with AI, legal, government relations, privacy, risk and compliance, public relations, executive leadership, etc.), are specifically trained to manage data breaches, and they undergo periodic exercises (e.g., table-top exercises) to prepare for and test their incident detection and response procedures.[32]

Organizations also have mechanisms in place to report unlawful breaches of personal data used in AI systems to regulators and individuals. Typically, organizations have guidelines to help determine if a given breach is a reportable one. While organizations will notify relevant parties of such breaches when necessary and required by law, they may also notify stakeholders voluntarily to maintain the trust they have built.

7.3 Responding to internal and external incidents and complaints

All the organizations included in our study are leveraging their existing channels of communication and offering various pathways to report, manage, and respond to both internal incidents of noncompliance and external consumer questions, complaints, and requests.

1. **Internal incident response**

   For internal incidents, many organizations want their employees to be aware of available lines of communication and use them to report incidents of misconduct or noncompliance.

   Some examples of such avenues include:

   - Ethics Hotline
   - Anonymous Incident Response Report
   - Open Q&A time during "all-hands" meeting with executive leadership
   - Escalation pathways to respond to emergent cases

2. **External redress mechanisms**

   Organizations use existing channels to allow external stakeholders, including business customers and everyday consumers, to file complaints, report issues, or ask questions regarding their rights as users of an AI service. Certain automated decision making (ADM) technologies are covered in several data protection laws,[33] so companies already have certain channels in place to comply with requests related to those laws. Organizations are leveraging those same channels for requests related to AI technologies to ensure human oversight and intervention. Some actions available to stakeholders through these channels include submitting a complaint about a decision; requesting a review of a decision;

32 See more about how accountable organizations manage data-related security breaches in CIPL's Report of the CIPL Accountability Mapping Project, available at https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_accountability_mapping_report__27_may_2020__v2.0.pdf

33 Some data protection laws that cover ADM technologies include the EU General Data Protection Law (GDPR), Brazil's General Data Protection Law (LGPD), the California Consumer Privacy Act (CCPA), and the US Equal Credit Opportunity Act (ECOA).

requesting information on how a decision was made; correcting or appealing a decision; or opting out of a business' use of ADM technology.

However, organizations expect forthcoming AI-focused regulations to require greater human oversight of ADM-driven AI systems, particularly those deemed high-risk. While some organizations have not yet fielded enough AI-related requests to warrant the creation of a unique communication channel, they are preparing to establish such channels in the near term.

CASE STUDY 37: Leveraging the customer assurance program (CAP) to denote high priority issues

An organization is using its pre-existing avenues to collect requests and complaints from customers. Each report is triaged for urgency. AI-related cases are classified as a "CAP case" and automatically marked as high-priority to address the issue as soon as possible.

Appendix A –Emerging Best Practices in Accountable AI Programs, Mapped to the CIPL Accountability Framework

The following table outlines a sample of emerging best practices and examples from accountable AI programs used by organizations from different sectors, geographies, and sizes. These practices are mapped to the corresponding element of the CIPL Accountability Framework. The practices are not intended to be mandatory industry standards but rather serve as examples of how companies are implementing specific practices to foster accountability in their development, deployment, and use of AI technologies. Each of the following should be calibrated based on risks, industry context, business model, size, and maturity level of the organization.

ACCOUNTABILITY ELEMENT	RELATED PRACTICES
Leadership and Oversight	• Establishing "tone from the top" and demonstrating a commitment to advance ethics, values, and specific principles in AI development, deployment, and use • Implementing systematic processes and escalation pathways for AI-related decision making • Establishing AI ethics oversight bodies or committees (internal or external) to review risky AI use cases and promote ongoing improvements to AI practices • Appointing a board member for AI oversight • Appointing a responsible AI lead, AI officer, or AI champion • Setting up an internal interdisciplinary AI board or AI committee • Establishing organization-wide AI ethics principles • Ensuring inclusion and diversity in AI model development and AI product teams • Creating a centralized governance framework with oversight from the top that still provides flexibility within internal teams • Expanding the remit of privacy teams to include AI-related responsibilities • Leveraging the expertise of other relevant teams (e.g., engineering, data science, legal, ethics and compliance, etc.) to ensure multidisciplinary, cross-functional AI teams • Encouraging employee reporting throughout all levels of the organization by offering escalation pathways to resolve potential AI-related issues
Risk Assessment	• Developing algorithmic impact assessments or fairness assessment tools to monitor and continuously test algorithms to avoid human bias, unfair discrimination, and "concept drift" throughout the entirety of the AI lifecycle • Requiring AI risk assessments at multiple points throughout the AI lifecycle, particularly for new or updated use cases or applications • Creating ethics, human rights, and/or data protection impact assessments • Creating a risk taxonomy that categorizes AI-related risks and allows for uniform assessment • Keeping a centralized repository of all risk assessment documentation • Developing standardized risk assessment methodologies that consider the benefits of the AI application or use, the likelihood and severity of risk factors on individuals and/or society, the level of human oversight needed for individually automated decisions with significant impact (e.g., legal ramifications), the ability to explain the technology in the appropriate context, and the ability to audit its effectiveness • Documenting considerations (e.g., accuracy, data minimization, security, transparency, scope of impact, benefits to society) for high-risk processing

ACCOUNTABILITY ELEMENT	RELATED PRACTICES
Risk Assessment	• Assessing data quality against key performance indicators (KPIs) • Evaluating the data vis-à-vis the purpose of its use (i.e., the quality of the data, its provenance, whether it's personal, synthetic, in-house, or externally sourced) • Developing frameworks for data preparation and model assessment – including feature engineering, cross-validation, back-testing, standardized KPIs • Enabling close collaboration between business and data experts (e.g., data analysts, data engineers, IT, and software engineers) on a regular basis to assess accuracy, ensure appropriate outputs, and allow for proper use of the model • Using privacy enhancing technologies (PETs) to preserve the privacy and security of AI systems • Outlining escalation pathways to send AI-related issues to an AI ethics council or other oversight body
Policies and Procedures	• Adopting specific AI policies and procedures on how to develop, deploy, or sell AI • Drafting policies on the application of privacy and security by design principles throughout the AI lifecycle • Setting rules on the level of verification for data input and output • Requiring pilot testing of AI models before release • Specifying the use of protected data (e.g., encrypted, pseudonymized, tokenized, or synthetic data) in training AI models • Creating a glossary of AI-related terms for internal use and reference • Promoting the use of smaller, higher quality data sets • Cleaning and curating data sets before model training through automated or manual checks • Considering relevant and appropriate use of PETs and PPTs to integrate privacy and security controls into AI models • Outlining special considerations for organizations creating and selling AI models, software, applications • Developing a fairness or AI impact assessment to analyze and mitigate AI-related risks • Creating due diligence/self-assessment checklists or tools for business partners deploying AI • Clearly defining escalation steps for reporting high-risk AI issues • Implementing an ideation phase with all stakeholders (e.g., data scientists, business, final user, control functions) where needs (including explainability), outcomes, validations rules, maintenance, and budget are discussed • Implementing specific policies for internal GenAI use • Requiring consideration for diversity in relevant teams and business functions • Implementing internal policies in parallel with forthcoming AI regulation • Translating internal principles-based policies to third-party vendor agreements, language, and due diligence processes • Creating processes for review of high-risk AI use cases by an AI ethics board or council
Transparency	• Tailoring transparency measures for the different needs of end users, regulators, business partners, and internal stakeholders at all stages of the AI lifecycle • Communicating disclosures in a simple, easy-to-understand manner • Considering how AI disclosures can be inclusive and accessible for those with special needs/disabilities • Establishing a transparency trail to explain automated decision-making and broad workings of algorithms • Providing notice when the system relies on AI/ML • Providing counterfactual information (e.g., how different inputs can affect the output of an AI model) • Understanding customers' expectations and deploying AI technologies based on their readiness to embrace AI • Implementing tiered transparency • Defining criteria for internal deployment of AI technologies based on usage scenarios and communicating them to users • Publishing model or system cards (i.e., short documents accompanying AI models that describe the context in which a given model is intended to be used and how the model performs in a variety of conditions) • Creating a data hub for information regarding data governance, data accessibility, data lineage, data modification, data quality, etc. • Tailoring transparency to the identified risk (e.g., using watermarking for generative AI output) where possible and appropriate • Participating in benchmarking opportunities, public engagement, and regulatory sandboxes • Using visualization tools to depict difficult, technically complex concepts to end users

Training and Awareness	• Providing specific training for data scientists and engineers, including how to address relevant ethical issues (e.g., how to limit and address bias) • Creating opportunities for cross-functional training (e.g., between privacy professionals and AI engineers) • Tailoring trainings regarding ethics and fairness in AI for relevant teams • Compiling and making available AI use case information where relevant risks have been mitigated or deployment has been halted • Creating a "translator" role that helps explain the impact and technical capacities and limitations of AI • Sharing case studies to help employees learn how to address potentially complex, ethically challenging AI cases • Incentivizing compliance with completing ethics training by pairing it with eligibility for bonuses, pay raises, and/or promotions, or incorporating it into other mandatory training activities
Monitoring and Verification	• Incorporating "human in the loop" (HITL) in design, oversight, and redress • Identifying and understanding which business functions are using AI • Providing the capability for human audit of input and output • Ensuring human review of individual decisions with legal or similarly significant effects • Monitoring the data ecosystem—from data flow in, through data process, to data flow out • Using different auditing techniques • Deploying counterfactual testing techniques • Pre-defining AI audit controls • Creating an internal audit team with expertise in AI and other emerging technologies • Allowing human control or intervention where technically possible and reasonably necessary • Monitoring AI models (e.g., back-testing and feedback loop) and conducting ongoing maintenance • Red teaming and adversarial testing of AI models
Response and Enforcement	• Enabling redress mechanisms to remedy an AI decision • Permitting redress through a human, not to a bot • Developing communication channels for internal (e.g., for employees) and external (e.g., end users, business customers) to report and address feedback, complaints, requests, etc.

About the Centre for Information Policy Leadership

CIPL is a global data privacy and cybersecurity think tank in the law firm of Hunton Andrews Kurth LLP and is financially supported by the law firm and 90 member companies that are leaders in key sectors of the global economy. CIPL's mission is to engage in thought leadership and develop best practices that ensure both effective privacy protections and the responsible use of personal information in the modern information age. CIPL's work facilitates constructive engagement between business leaders, privacy and security professionals, regulators and policymakers around the world. For more information, please see CIPL's website at

http://www.informationpolicycentre.com/.

DC Office
2200 Pennsylvania Avenue
Washington, DC 20037
+1 202 955 1563

London Office
30 St Mary Axe
London EC3A 8EP
+44 20 7220 5700

Brussels Office
Avenue des Arts 47-49
1000 Brussels
+32 2 643 58 00

24085.0224