



**EQUAL**<sup>TM</sup>

---

**An Insider's Guide to Designing  
and Operationalizing a Responsible  
AI Governance Framework**

September 2023

# CO-AUTHORS

---

**Laxmi Anish**, *PepsiCo, Inc.*

**Kathy Baxter**, *Salesforce Inc.*

**Dawn Bloxwich**, *Google DeepMind*

**Catherine Goetz**, *LivePerson*

**Tina Huang**, *EqualAI*

**Gabrielle Kohlmeier**, *Verizon*

**David Lincoln**, *Verizon*

**David Marcos**, *Microsoft Corporation*

**Amanda Muller**, *Northrop Grumman Corporation*

**Mike Tang**, *Verizon*

**Reggie Townsend**, *SAS Institute*

**Miriam Vogel**, *EqualAI*

**Diya Wynn**, *Amazon Web Services*

# TABLE OF CONTENTS

---

- I. Executive Summary** 5
- II. Introduction** 9
- III. Background on the EqualAI Responsible AI Summit** 10
- IV. Responsible AI Governance Framework** 12
  - A. First: Ground the Process in Your Reality 12
  - B. Second: Identify Values 12
    - 1. Trust 13
    - 2. Culture 14
    - 3. Accountability 15
    - 4. Multistakeholder Engagement 16
  - C. Third: Determine Responsible AI Principles 17
  - D. Fourth: Establish Accountability and Clear Lines of Responsibility 21
  - E. Fifth: Use Tools to Support AI Governance 24
    - 1. Documentation 24
    - 2. Defined Process 24
    - 3. Multistakeholder Review 26
    - 4. Metrics, Monitoring, and Reevaluation 26
- V. Further Discussion and Open Questions** 28
- VI. Conclusion** 31

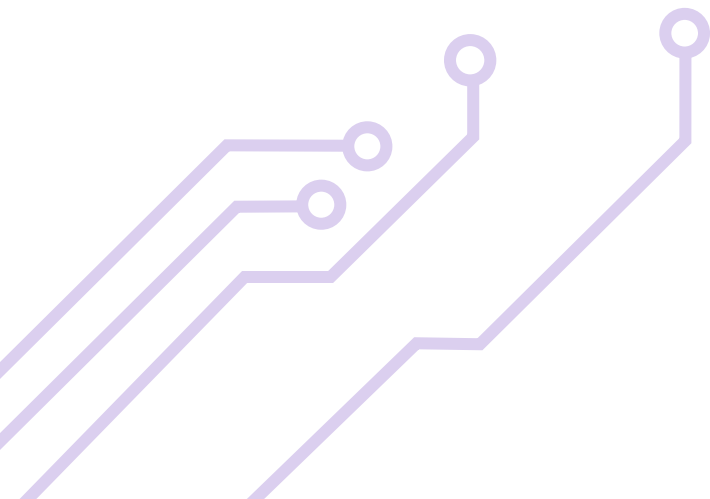
# ACKNOWLEDGMENTS

---

We are grateful for all those who played a crucial role in the production of this white paper and our EqualAI Badge Program. We would like to recognize the esteemed faculty for our Badge Program, including Kathy Baxter, Meredith Broussard, Andrew Burt, Natasha Crampton, Jen Gennai, Cathy O’Neil, and Reva Schwartz, and the distinguished speakers at the EqualAI Responsible AI Summit, including Rep. Don Beyer (VA-8), Dr. Seth Center, Rep. Ted Lieu (CA-36), and Rep. Jay Obernolte (CA-23) for contributing to informative discussions and insights that shaped the findings and perspectives of our conversations and ultimately our paper, as well as inspiring confidence in our policymakers to navigate these complex issues.

We would also like to extend sincere thanks to our invaluable EqualAI board members and advisers, including Jonathan Dotan, Victoria Espinel, and Reggie Townsend for their involvement at the Summit. The productive conversations that took place were made possible in large part by the elegant, welcoming venue and the staff at the House at 1229, including Kathleen Buhle, Johanna Harris, and Karima Ouazzani; we are grateful for their unparalleled professionalism while hosting the Summit. Special thanks are also due to the NP Agency team—Alexandria Dissell, Tom McMahon, and Cara Morris Stern—for editing, designing, and promoting the white paper as well as our indispensable EqualAI community, including James Beasley, Kristi Boyd, Hannah Dudley, Becca Kahn, and Meghna Sinha, for their feedback and contributions.

We would like to express our gratitude for our copy editor, Nancy King, and EqualAI staff—James Bond and Jim Wiley—for their work in bringing the white paper across the finish line. Lastly, we would like to recognize Microsoft Corporation for their significant contributions to the white paper, including their instruction and participation in the Badge Program, and support for the Responsible AI Summit.



# EXECUTIVE SUMMARY

---

As businesses—and individuals—increase their dependence on and investment in artificial intelligence (AI)<sup>1</sup>, there is a critical need to establish and align on best practices that promote responsible AI to earn and maintain trust in these systems. Since 2017, AI adoption has doubled, and organizations that have embraced AI are experiencing the highest financial returns relative to competitors.<sup>2</sup> Additionally, numerous studies report that consumers overwhelmingly expect businesses to be responsible and ethical when adopting and developing AI technology.<sup>3</sup> Now more than ever, organizations<sup>4</sup> must understand and implement practices to ensure their AI is *responsible*—by which we mean safe, inclusive, and effective for all possible end users. An organization must not only consider its consumers but also the impact these new technologies will have on their employees and the company’s culture, as well as any unintended use cases.

Currently, there is a lack of national, let alone global, consensus on standards for responsible AI governance. While there is some indication of progress<sup>5</sup> to come on this front, it is not imminent, and organizations cannot wait for the regulatory and litigation landscape to settle before adopting best practices for AI governance. The potential harm and liability associated with the complex AI systems currently being built, acquired, and integrated is too significant to delay adoption of safety standards. **To help companies prepare for this reality, EqualAI convenes leaders across industry, government, and civil society to align on risks, liabilities, and best practices in establishing and operationalizing responsible AI governance.**

---

<sup>1</sup> EqualAI relies on the [NIST AI Risk Management Framework’s](#) definition of an AI system: “An AI system is an engineered or machine-based system that can, for a given set of objectives, generate outputs such as predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy.” (Adapted from: OECD Recommendation on AI:2019; ISO/IEC 22989:2022.)

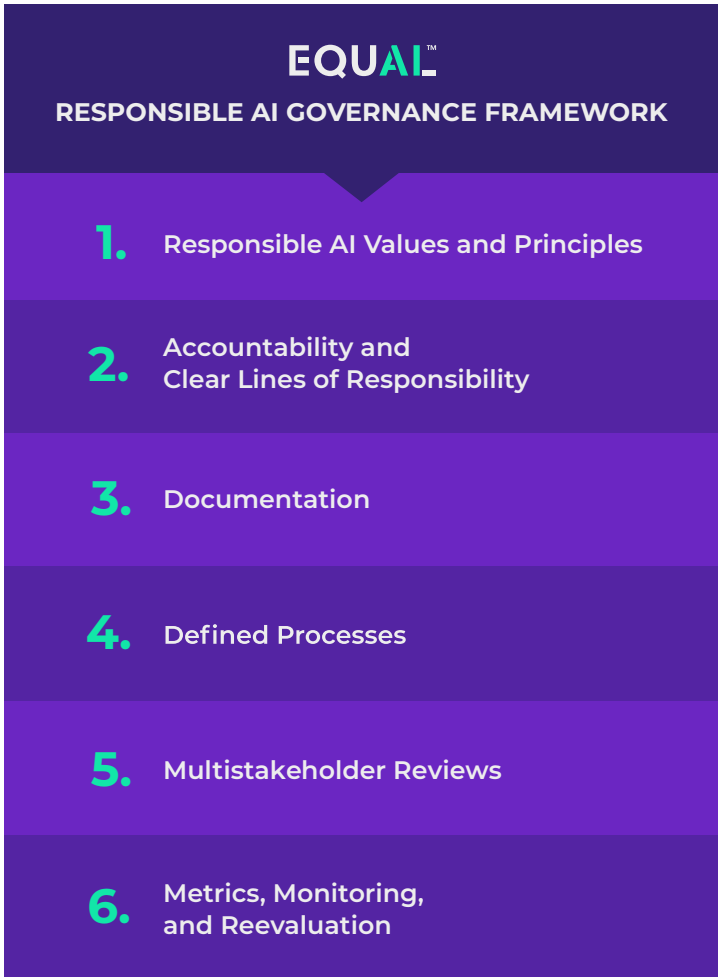
<sup>2</sup> McKinsey & Company, “The State of AI in 2022—and a Half Decade in Review,” McKinsey & Company, December 6, 2022, <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2022-and-a-half-decade-in-review>.

<sup>3</sup> Business Wire, “WE Communications’ Brands in Motion 2018 Global Study: Escalating Consumer Expectations Push Brands to Deliver on Innovation, Ethical Responsibility and Functionality,” September 12, 2018, <https://www.businesswire.com/news/home/20180912005040/en/Communications%E2%80%99-Brands-Motion-2018-Global-Study-Escalating>; Cisco, “Ninety-Two Percent of Organizations Think They Need to Do More to Assure Customers about How Their Data Is Used in AI, New Cisco Research finds,” January 24, 2023, <https://investor.cisco.com/news/news-details/2023/Ninety-two-percent-of-organizations-think-they-need-to-do-more-to-reassure-customers-about-how-their-data-is-used-in-AI-new-Cisco-research-finds/default.aspx>.

<sup>4</sup> Note that we use “organization” and “company” interchangeably in this paper, given that our intended audience is both types of entities and the findings presented here are equally applicable.

<sup>5</sup> “The EU Artificial Intelligence Act,” June 14, 2023, <https://www.artificial-intelligence-act.com/>; National Institute for Standards and Technology, “Artificial Intelligence Risk Management Framework,” January 2023, <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>.

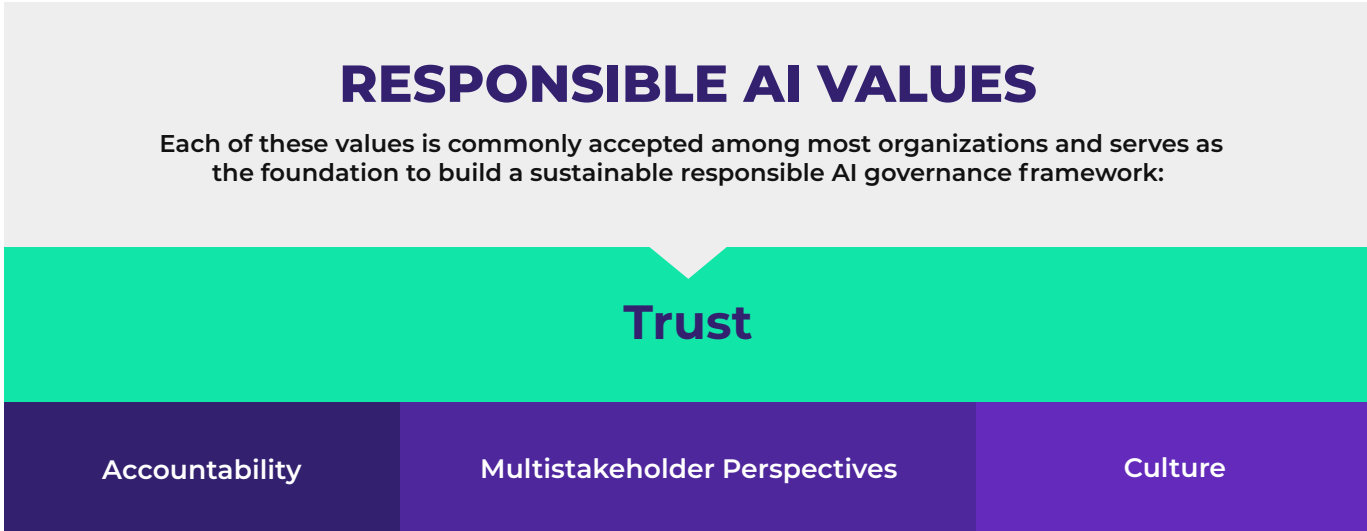
**The primary goal of this white paper is to serve as a resource for organizations of any size, sector, or maturity that are adopting, developing, using, and implementing AI systems with an internal and public commitment to do so responsibly. We hope this will be helpful to additional audiences, such as policy leaders who want to understand how industry leaders who are committed to this effort are taking the initiative to ensure responsible AI governance.**



This white paper builds on the discussion in the culminating seventh session of our [EqualAI Badge Program](#), where senior executives gather to address best practices in responsible AI governance. The final framework they aligned on consists of six main pillars.

Each of these six pillars plays a critical role in establishing the groundwork for an organization’s responsible AI governance framework. Together, they empower executives and employees to leverage resources, tools, and guidance to proactively ensure AI is adopted, developed, used, and implemented consistently, safely, and inclusively. Implementation of this responsible AI governance framework will also signal to consumers and the general public that an organization is committed to earning and maintaining its trust.

At the initial stages of building an AI framework, our participants agreed on key AI values and principles that serve as the foundation to develop a cohesive process and inform end goals. These include:



# RESPONSIBLE AI PRINCIPLES

Each of these principles is important to the effective operations of an organization but should be considered as they apply to AI systems more specifically. They include:

- 1. Preservation of Privacy:** *Assure privacy and protection of data and its subjects.*
- 2. Transparency:** *Communicate values, principles, framework, policies, and decision-making process on AI, both internally and externally.*
- 3. Human-Centered Focus:** *Commit to building and using AI that benefits human life and amplifies and augments, rather than displaces human abilities.*
- 4. Respect for Individual Rights and Societal Good:** *Demonstrate respect for human and civil rights and systems that are built to promote social good.*
- 5. Open Innovation:** *Commit to innovation that drives openness and collective sharing.*
- 6. Rewarding Robustness:** *Build a reward system that prioritizes technical robustness.*
- 7. Continuous Iteration and Review:** *Establish continuous learning loops to integrate feedback from all stakeholders, users, and those impacted downstream.*
- 8. Enlist Employees' Involvement:** *Enlist employees as your front line to promote responsible AI and to ensure they feel safe and empowered to flag potential concerns.*
- 9. Prioritize Fairness Through Accountability:** *Promote fair and unbiased AI systems through evaluations of data, algorithms, and humans.*
- 10. Human-in-the-Loop:** *Incorporate human input and oversight into all stages of AI decision-making.*
- 11. Professional Development:** *Invest in education and training to upskill and/or re-skill employees.*

With a grounding in these values and principles, an organization can begin to set the foundation of its responsible AI governance framework. This white paper explores the six pillars, explains their importance, and offers methods for organizations to operationalize an AI framework consistent with its unique purposes, goals, values, and principles.



# INTRODUCTION

---

EqualAI works with organizations that are eager to seize the benefits of AI while mitigating potential harms and biases. In recent years, companies, governments, and nongovernment entities have progressively developed AI principles and governance frameworks.<sup>6</sup> This is an important development as organizations are increasingly becoming “AI companies,” meaning those that use AI in pivotal areas ranging from hiring and other HR functions to financial or benefits-related determinations.

In May 2023, EqualAI gathered industry leaders to align on responsible AI principles to create a governance framework applicable to any organization regardless of industry, sector, size, or maturity. We are often asked to share our lessons and discussions on best practices with the broader community. We very much agree that sharing this information is in itself a best practice, both to heighten awareness of and alignment on best practices, as well as to crowdsource these evolving standards and provide an opportunity to gather broader input. **This white paper shares both themes and learnings from our discussion at the Responsible AI Summit as well as lessons discussed and operationalized throughout the Badge Program.**

**Intent and audience:** We are excited to work with our partners (noted herein) to create what is arguably a first-of-its-kind document, which incorporates the views of multiple companies with multidisciplinary perspectives on best practices in responsible AI governance, and we are grateful to our participants and partners for helping us to create this tool. The intended audience includes organizations that are building or deploying responsible AI governance and policymakers looking to understand the challenges encountered and practices adhered to by those companies leading the field of responsible AI governance as they consider the creation of appropriate guardrails.

**A road map for this document:** This white paper begins with a background on the EqualAI Responsible AI Summit, including information about the EqualAI Badge Program; the Summit’s goals; and participating companies. The paper then offers a responsible AI governance framework, which includes commonly accepted organizational values and explains how they relate to responsible AI, key AI principles for organizations to consider adopting, and tools to help implement and monitor responsible AI governance efforts. Finally, the white paper dives into highlights of the participants’ deliberations, noting questions for which they did not reach consensus or have time to consider.

**The end goal:** This white paper intends to guide organizations with top-line best practices used by leading practitioners and thought leaders in responsible AI governance, as they adopt, develop, use, and implement AI responsibly.

---

<sup>6</sup> Mina Narayanan and Christian Schoeberl, “A Matrix for Selecting Responsible AI Frameworks,” Center for Security and Emerging Technology, June 2023, <https://cset.georgetown.edu/publication/a-matrix-for-selecting-responsible-ai-frameworks/>.

# BACKGROUND ON THE EQUALAI RESPONSIBLE AI SUMMIT

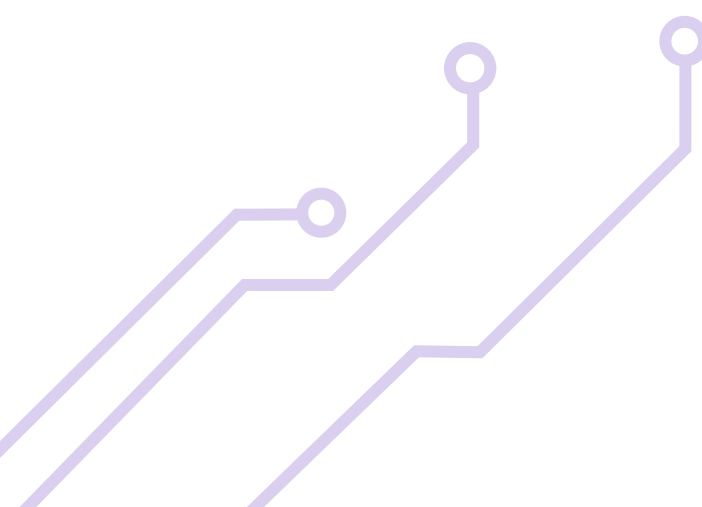
---

The EqualAI Responsible AI Summit is a culmination of the **EqualAI Badge Program**, where past and present Badge Program participants gather to align on responsible AI principles and best practices with the goal of developing consensus on a responsible AI governance framework. Our participants recognized that the needs of an organization will vary depending on its size, sector, maturity, resources, level of AI use and development, and other critical factors. As such, they focused on identifying principles and creating a framework that includes transferable characteristics regardless of these variables and offering different options for organizations to customize responsible AI efforts. You will find the high-level themes and findings from our discussions in this paper.

## OVERVIEW OF EQUALAI

EqualAI is a nonprofit organization leading the movement to reduce unconscious bias and other harms in the development and deployment of artificial intelligence. We work with leaders and experts across business, technology, and government to develop ethical standards, technical tools, and legislative solutions for responsible AI governance.

Our flagship programs include the Badge Program for corporate leaders as described above, as well as our CA-accredited Continuing Legal Education (CLE) course, designed to help lawyers understand their role in reducing bias in AI and, in turn, better serve their clients. In addition, EqualAI co-hosts a podcast, ***In AI We Trust?***, featuring AI thought leaders dedicated to defining, developing, and deploying responsible, trustworthy AI.



# Context: What is the EqualAI Badge Program?

The EqualAI Badge Program prepares senior executives at companies developing or using AI systems to govern AI in ways that will help reduce potential harms and liability, empower employees to identify AI risks or risks associated with developing and deploying AI technologies, and broaden the potential use cases and consumer base. Participants stem from a broad cross section of industry (e.g., cloud-based software, service providers, sales, etc.) and roles (chief data officers, general counsel, chief privacy officers, chief responsible AI officers, etc.).

In this program, responsible AI governance is divided into six vantage points. Each session brings AI experts to meet with the Badge community and align on best practices for responsible AI governance. Interactive monthly panel discussions help inform participants on how to operationalize AI principles with the goal of creating more inclusive AI systems and reducing liability. Senior executives learn best practices and become part of a community of leading experts and executives in the responsible AI field. Upon completion, Badge alumni are invited to quarterly convenings to continue their education and engagement with the EqualAI community.

## Badge Program topics and speakers on monthly panel sessions generally include (with some variation):

- **How Bias Translates and Embeds in Our AI** with Meredith Broussard (New York University);
- **Ethical Matrix for Creating AI** with Cathy O'Neil (ORCAA);
- **Is Your AI Safe to Launch** with Kathy Baxter (Salesforce);
- **Tools and Strategies to Operationalize Responsible AI Governance** with Jen Gennai (Google) and Natasha Crampton (Microsoft);
- **The Role of Lawyers in Addressing Bias in AI** with Alexandra Reeve Givens (Center for Democracy and Technology), Andrew Burt (bnh.ai), and Tom Lue (DeepMind);
- **The AI Policy Landscape** with staff from Capitol Hill, U.S. Equal Employment Opportunity Commission (EEOC), and the National Institute of Standards and Technology (NIST)

## 2023 Spring Summit participants include senior executives from the following companies:

- Amazon Web Services (AWS)
- Verizon
- LivePerson
- Hewlett Packard Enterprise (HPE)
- Northrop Grumman
- Microsoft
- Salesforce
- PepsiCo
- DeepMind
- SAS Institute

# RESPONSIBLE AI GOVERNANCE FRAMEWORK

---

Establishing a responsible AI governance framework is crucial for any organization planning to design or integrate AI into their internal functions or external products and services. Such a framework helps to operationalize an organization's AI principles and values and offers a set of guidelines and best practices to ensure the intentional, transparent, and consistent development, deployment, and use of AI technologies. In addition, the framework will empower employees and allow an organization to engender trust in its AI use more specifically and its operations and intentions more generally.

## **A. GROUND THE PROCESS IN YOUR REALITY**

As an organization's leadership begins its responsible AI governance journey, it is helpful to understand how the company is currently using AI, and what plans teams have to integrate AI for pivotal functions. After surveying their particular AI landscape and horizon, it is time to develop AI principles that align with the organization's values and establish an infrastructure and process, as detailed below, to support these values and ensure they are not impeded by AI use. From there, leadership should ground the processes they develop with tangible use cases to pressure-test the new AI principles and customize a responsible AI governance framework suited to their specific needs. The process should include opportunities to review and iterate as the AI systems will continue to evolve and iterate.

## **B. IDENTIFY VALUES**

Responsible AI governance begins with an organization's stated or established values. Values are standards or qualities that set behavioral norms and serve as an organization's cultural cornerstone. They underlie business functions—from operations to strategy—that an organization implements to meet its performance goals, especially during dynamic periods of growth or change.<sup>7</sup> In light of rapidly advancing AI systems that have transformed and will continue to transform nearly every industry, ensuring an organization's values align with AI development and use is essential.

---

<sup>7</sup> Brent Gleeson, "Why Core Values Matter (and How to Get Your Team Excited about Them)," *Forbes*, March 30, 2021, <https://www.forbes.com/sites/brentgleeson/2021/03/30/why-core-values-matter-and-how-to-get-your-team-excited-about-them/?sh=e-5613b14afde>; John Coleman, "It's Time to Take a Fresh Look at Your Company's Values," *Harvard Business Review*, March 28, 2022, <https://hbr.org/2022/03/its-time-to-take-a-fresh-look-at-your-companys-values>.

Our Badge participants found that the following values are commonly adopted by most organizations leading in this space. The discussion below (not in any particular order) details why each value is important and how each can establish a foundation for responsible AI principles and governance.<sup>8</sup>

## ► **Trust**

An organization's success heavily depends on trust both internally, among employees, and externally, with consumers and stakeholders. Our Badge participants noted that without establishing trust in the organization's commitment to responsible AI governance, the process would be extremely challenging if not impossible to achieve. When employees do not have trust in their organization to operate fairly and with integrity, there are numerous, significant consequences: performance suffers, retention decreases, and, as a result, customers may lose trust and are less inclined to purchase products or retain services.<sup>9</sup>

A recent **Harvard Business Review** survey<sup>10</sup> found four core components that help establish trust with customers: data protection and cybersecurity, treating employees well, ethical business practices, and admitting mistakes quickly and honestly. The same survey revealed that business leaders highly value how an organization manages its value chains, deploys responsible AI, and reports on environmental, social, and governance (ESG) actions. These elements are key to developing an ecosystem of trust, which establishes a foundation for responsible AI governance.

Organizations should also make efforts to cultivate trust in AI systems. More trustworthy AI systems will, in turn, strengthen trust with employees and consumers. To this end, National Institute of Standards and Technology (NIST) has identified the following characteristics that make an AI system trustworthy<sup>11</sup>: accuracy, explainability, interpretability, privacy, reliability, robustness, safety, and security resilience. To promote the development of trustworthy AI, employees should feel empowered to document these characteristics and discuss shortcomings or improvements without negative recourse and, optimally, they should expect employers to show interest in suggested improvements.

---

<sup>8</sup> National Institute for Standards and Technology, "Artificial Intelligence Risk Management Framework."

<sup>9</sup> PwC, "Trust: The New Currency for Business," June 2022, <https://www.pwc.com/us/en/services/consulting/library/consumer-intelligence-series/trust-new-business-currency.html>; Shelley Smith, "Lack of Trust Can Make Workplaces Sick and Dysfunctional," *Forbes*, October 24, 2019, <https://www.forbes.com/sites/forbescoachescouncil/2019/10/24/lack-of-trust-can-make-workplaces-sick-and-dysfunctional/?sh=6ad7cc4444d1>; Susan Ladika, "Trust Has Never Been More Important," *SHRM*, July 31, 2021, <https://www.shrm.org/hr-today/news/all-things-work/pages/trust-has-never-been-more-important.aspx>.

<sup>10</sup> Tim Ryan, "How Business Can Build and Maintain Trust," *Harvard Business Review*, February 7, 2022, <https://hbr.org/2022/02/how-business-can-build-and-maintain-trust?registration=success>.

<sup>11</sup> Reva Schwartz et al., "Towards a Standard for Identifying and Managing Bias in Artificial Intelligence," National Institute for Standards and Technology, Special Publication 1270, March 2022, <https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1270.pdf>.

## ► Culture

Culture has been defined as the ways individuals behave, including the attitudes and beliefs that inform these behaviors, which can be explicitly stated or implicitly known as the norms governing how people should work or interact.<sup>12</sup> Building a culture around AI that prioritizes responsible governance is critical to ensuring employees feel encouraged, safe, and even rewarded for openly discussing challenges they encounter in their day-to-day work.<sup>13</sup> Further, fostering an inclusive culture will help organizations navigate the reality that not all AI risks are the same for all users and will help inform whether a given AI system is the appropriate or the best solution for a challenge.<sup>14</sup>

**Salesforce’s Ethical AI Maturity Model, which is discussed in Session 3 of the Badge Program, provides that organizations establish AI strategies that ultimately build a responsible AI culture. The model includes four stages with a bottom-up approach:**

1. Stage one involves individual advocates generating small-scale strategies and earning buy-in.
2. Stage two focuses on formal teams and resources to align efforts toward an executable strategic vision.
3. Stage three establishes new teams to develop measures and a long-term mentality for sustainable practices.
4. Stage four emphasizes an optimized and innovative approach to integrate ethics throughout the entire organization.<sup>15</sup>

A common theme discussed and promoted in the Badge Program is that every employee should feel as though they are on the front lines of detecting potential harms, and encouraged and rewarded for promoting AI safety and, thus, building trust.

---

<sup>12</sup> Denise Lee Yohn, “Company Culture Is Everyone’s Responsibility,” *Harvard Business Review*, February 8, 2021, <https://hbr.org/2021/02/company-culture-is-everyones-responsibility>.

<sup>13</sup> Schwartz et al., “Towards a Standard for Identifying and Managing Bias in Artificial Intelligence.”

<sup>14</sup> National Institute for Standards and Technology, “Artificial Intelligence Risk Management Framework,” January 2023, <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>.

<sup>15</sup> Salesforce, “Salesforce Debuts AI Ethics Mode: How Ethical Practices Further Responsible Artificial Intelligence,” September 2, 2021, <https://www.salesforce.com/news/stories/salesforce-debuts-ai-ethics-model-how-ethical-practices-further-responsible-artificial-intelligence/>.

Our Badge participants noted that Key Performance Indicators (KPIs) based only on team or employee “success,” as opposed to broader metrics, may skim over the real causes of system failures. Instead, companies should empower, encourage, and incentivize employees to contextualize mistakes with supervisors or through determined appropriate channels. Further, participants emphasized the importance of risk mitigation instead of risk avoidance.<sup>16</sup> This is because attempting to reach a “zero risk” policy may in practice be counterproductive, and virtually unattainable, since not all incidents and potential failures can be eliminated.<sup>17</sup>

Establishing a culture of responsible AI governance requires an investment of time and resources. This is a priority that must be underscored when establishing budgets and employment evaluations, including the hiring and support for those doing this work within an organization. This work also requires the inclusion of time in the AI integration and/or AI product development process to allow for assessment and mitigation, when necessary.

Further, setting up performance recognition, pay, and promotion incentives that foster trust based on the elements discussed in this section helps to ensure support for and adoption of AI risk mitigation efforts and to strengthen a responsible AI culture.<sup>18</sup>

## ► **Accountability**

Accountability in an organization means that individuals understand what is expected of them, can exercise agency or authority when appropriate, and take responsibility for delivering results.<sup>19</sup> Holding people or groups accountable for responsible AI governance is critical to ensuring proactive bias and harm mitigation.

To build meaningful accountability for responsible AI governance, an organization should start by defining clear lines of authority and designating individuals tasked with enforcing responsible AI governance (“responsible AI leaders”). Traditional KPIs may need to be adjusted when they are based only on team or employee “success,” as opposed to broader metric measures that take into account the reality of why an employee or team was unable to achieve their goals, such as identification of the project’s potential harms or damaging impacts.<sup>20</sup>

---

<sup>16</sup> Schwartz et al., “Towards a Standard for Identifying and Managing Bias in Artificial Intelligence.”

<sup>17</sup> National Institute for Standards and Technology, “Artificial Intelligence Risk Management Framework.”

<sup>18</sup> Schwartz et al., “Towards a Standard for Identifying and Managing Bias in Artificial Intelligence.”

<sup>19</sup> Michael Bazigos, Diana Ellsworth, and Drew Goldstein, “Where Accountability Really Matters,” McKinsey & Company, April 2016, [https://www.mckinsey.com/-/media/mckinsey/industries/travel%20logistics%20and%20infrastructure/our%20insights/where%20accountability%20really%20matters/where\\_accountability\\_really\\_matters.pdf](https://www.mckinsey.com/-/media/mckinsey/industries/travel%20logistics%20and%20infrastructure/our%20insights/where%20accountability%20really%20matters/where_accountability_really_matters.pdf).

<sup>20</sup> Ron Carucci, “It’s Time to Overhaul Our Understanding of Accountability,” *Forbes*, June 4, 2022, <https://www.forbes.com/sites/roncarucci/2022/06/04/its-time-to-overhaul-our-understanding-of-accountability/?sh=3e3321883497>.

Responsible AI leaders must clearly communicate expectations while empowering and equipping employees with the appropriate resources to meet these expectations. If expectations are unmet, organizations can take the opportunity to promote transparent conversations to learn from mistakes and understand what various individuals and/or divisions could have done differently at any point during the AI lifecycle.<sup>21</sup> This is especially important due to the evolving, and oftentimes unpredictable, nature of AI systems. Accountability should also be embedded within and across teams—such as sales, marketing, legal, DEI, and others—that are directly or indirectly involved with training, developing, deploying, using, or monitoring AI systems.<sup>22</sup> Finally, organizations should identify rewards to reinforce behaviors that align with upholding responsible AI values and principles.

### ► **Multistakeholder Engagement**

As organizations begin to integrate AI into internal processes or client-facing functions, collaborating with or undergoing review with relevant stakeholders is critical to ensuring comprehensive decision-making on AI. In determining the appropriate representatives to participate in multistakeholder reviews, organizations should take into account a variety of potential use cases and impact on users, especially those in underrepresented and/or marginalized communities.<sup>23</sup> In particular, organizations should establish processes to solicit insights from underrepresented and marginalized communities that could be impacted by the AI system initially or downstream in order to identify conceivable use cases and potential risks that may not be readily apparent otherwise. Ensuring a diversity of perspectives and efficient collaboration with both internal and external stakeholders in AI development and deployment enables employees to drive AI innovation, problem-solve, and adapt its development to quickly changing environments.<sup>24</sup>

---

<sup>21</sup> Ibid.

<sup>22</sup> Schwartz et al., "Towards a Standard for Identifying and Managing Bias in Artificial Intelligence."

<sup>23</sup> Michael Li, "To Build Less-Biased AI, Hire a More-Diverse Team," *Harvard Business Review*, October 26, 2020, <https://hbr.org/2020/10/to-build-less-biased-ai-hire-a-more-diverse-team>; Arun Shastri, "Diverse Teams Build Better AI. Here's Why," *Forbes*, July 1, 2020, <https://www.forbes.com/sites/arunshastri/2020/07/01/diverse-teams-build-better-ai-heres-why/?sh=7ecab69f77b3>; Esther Ajao, "Diversity Within Your AI Team Can Reduce Bias," *TechTarget*, December 9, 2022, <https://www.techtarget.com/searchenterpriseai/news/252528214/Diversity-within-your-AI-team-can-reduce-bias>; Zachariah Chou, "Diverse AI Teams Are Key to Reducing Bias," *VentureBeat*, July 22, 2021, <https://venturebeat.com/ai/diverse-ai-teams-are-key-to-reducing-bias/>.

<sup>24</sup> McKinsey & Company, "Leading Off," April 11, 2022, <https://www.mckinsey.com/-/media/mckinsey/email/leadingoff/2022/04/11/2022-04-11b.html>.



In addition, identifying a perspective or role that is missing in the AI planning and design discussion is an important point that should be considered at each stage of the AI lifecycle. Given the changing use cases and continuous iterations of AI, it is equally important to always ask who else needs to be at the proverbial “table” or, as Cathy O’Neil, EqualAI senior adviser, acclaimed author, and instructor in our Badge Program, puts it: “for whom could this technology fail?” This work is crucial to ensuring responsible AI development and use.<sup>25</sup>



## C. DETERMINE RESPONSIBLE AI PRINCIPLES

Aligning on AI principles offers a pathway for organizations to operationalize their values by setting rules and standards to guide decision-making related to AI development and use. These principles are public statements of how an organization intends to operate in a landscape of transformative technologies.

To this end, establishing AI principles is a key step to set the groundwork for a responsible AI governance framework. Though applicability and prioritization may vary among different industries, our participants identified the following principles they consider to be critical regardless of the industry or organization’s size. Below are key AI principles, not listed in a particular order, that our Badge community identified. Each principle is labeled with the responsible AI value(s) with which it aligns. Depending on applicability and feasibility, organizations should aim to adopt as many as possible, if not all, of the AI principles.

<sup>25</sup> Roger Burkhardt, Nicolas Hohn, and Chris Wigley, “Leading Your Organization to Responsible AI,” QuantumBlack AI by McKinsey, May 2, 2019, <https://www.mckinsey.com/capabilities/quantumblack/our-insights/leading-your-organization-to-responsible-ai>; Li, “To Build Less-Biased AI, Hire a More-Diverse Team.”

## **1. Preservation of Privacy: Assure privacy and protection of data and its subjects (Trust, Accountability)**

If not already in place, organizations should establish generally applicable policies that assure the privacy and protection of data and its owners and assure they address data issues relating to AI. With regard to privacy, given its legal foundation, a key question is whether to adhere to existing requirements, such as the European Union General Data Protection Regulation (GDPR), or to provide additional protections. It is also important for leaders to recognize that not all data owners are subjects of that data. For example, a third-party data vendor may track a consumer's shopping habits and own the data but is not the subject of the data itself. Consequently, organizations should thoroughly understand the privacy policies of entities they are acquiring from and evaluate whether those policies align with their own organization's responsible AI governance principles and framework. Further, these policies should outline a clear process for addressing breaches in privacy.

## **2. Transparency: Communicate values, principles, framework, policies, and decision-making process on AI, both internally and externally (Trust, Multistakeholder Engagement, Culture)**

An organization's responsible AI values, principles, and governance framework should be made available and accessible for internal and external stakeholders. While these documents may look different for each audience, they should be digestible for a broad variety of stakeholders with key terms clearly defined. Each document should also include a point of contact for individuals to reach out to with questions or concerns.

Where and when possible, organizations should explain their decision-making processes around responsible AI issues to stakeholders, employees, and consumers. These explanations should be easily understood by technical and non-technical teams, using language designed to communicate with foreseeable audiences, such as consumers or end users who operate in different languages or at different reading levels. Such transparency will help build trust internally and externally, while also mitigating misconceptions or confusion around an organization's approaches, priorities, and positions on AI.

## **3. Human-Centered Focus: Commit to building and using AI that benefits human life and amplifies and augments, rather than displaces human abilities (Trust, Multistakeholder Engagement)**

In some use cases and particularly with elevated risk use-cases, AI should be used to augment, not replace, human capabilities.<sup>26</sup> At every point in the AI lifecycle, decision-makers or developers must consider how the technology will directly

---

<sup>26</sup> National Institute for Standards and Technology, "Trustworthy & Responsible AI Resource Center," [https://airc.nist.gov/AI\\_RMF\\_Knowledge\\_Base/Playbook/Govern#Govern%2013](https://airc.nist.gov/AI_RMF_Knowledge_Base/Playbook/Govern#Govern%2013).

or indirectly impact humans. Decisions should then be made with consideration of how the technology will have a net-positive impact on humanity. In addition, employees should be encouraged to refrain from using AI to fulfill their roles or certain tasks without approval or oversight. They must also maintain due diligence to ensure AI-enabled decisions are fair and effective for various audiences, particularly underrepresented and minority groups.

**4. Respect for Individual Rights and Societal Good: Demonstrate respect for human and civil rights and systems that are built to promote social good (Trust, Culture)**

At every point in the AI lifecycle, employees should consider the implications the model may have on human and civil rights. Organizations should also be transparent with clients and stakeholders about how AI could potentially impact human and civil rights, both positively and negatively, and how the organization plans to either leverage or mitigate those implications, respectively.

**5. Open Innovation: Commit to innovation that drives openness and collective sharing (Multistakeholder Engagement, Culture)**

Organizations should establish norms that promote openness within the organization and collective sharing of new ideas, practices, setbacks, and breakthroughs. Collective wins should be celebrated, and failures should be viewed as learning opportunities for all. By building a safe and open culture, organizations can best position themselves to detect potential harms and biases early and to proactively address them before they escalate into larger issues.

**6. Rewarding Robustness: Build a reward system that prioritizes technical robustness (Culture, Trust)**

Organizations should ensure employees understand that working toward “technical robustness,” whereby AI systems reliably perform as they are intended to, is a top priority. To this end, AI development timelines should be considered early in the development lifecycle account to ensure time is built in for technical robustness.

Employees should also know they will not suffer negative consequences if honing technical robustness may delay the production timeline. Further, organizations should encourage employees to highlight areas of brittleness, in which an AI system fails to perform as intended, and create a process to address these concerns. Overall, employees should be rewarded for flagging system errors, failures, or brittleness, as well as proposing helpful resolutions or delaying a production timeline to improve robustness.

**7. Continuous Iteration and Review: Establish continuous learning loops to integrate feedback from all stakeholders, users, and those impacted downstream (Multistakeholder Engagement, Culture)**

Internal and external stakeholders should have an opportunity to provide feedback on an organization’s principles and responsible AI governance framework.

There should be a clear process in which they can engage with the organization to deliver feedback, and have the option to remain anonymous. Such feedback loops allow the organization to continuously learn and iterate on their principles and framework to meet the evolving needs and realities of responsible AI governance.

**8. Enlist Employees Involvement: Enlist employees as your front line to promote responsible AI and to ensure they will feel safe and empowered to flag potential concerns (Culture, Multistakeholder Engagement)**

Employees play a critical role in ensuring responsible AI governance as they are often at the front lines of developing, deploying, using, and monitoring AI systems. Organizations should establish mechanism(s) for employees to alert leadership to potential concerns they have either witnessed or experienced, with the option to remain anonymous. Further, leadership should initiate regular conversations with employees about responsible AI and why it is important for the organization, and empower employees to proactively act on responsible AI principles.

**9. Prioritize Fairness Through Accountability: Promote fair and unbiased AI systems through evaluations of data, algorithms, and humans (Trust, Accountability)**

Holding employees and leaders accountable for fulfilling their duties for responsible AI governance is key to ensuring sustainability. Organizations should define and set metrics that align with their values and goals for mitigating bias and promoting fairness. This includes addressing biases that may emerge at any point in the AI lifecycle, and ensuring the technology does not discriminate against individuals or groups based on any protected, or non-protected, characteristics. Leaders should also establish an accountability process to identify the cause of potential risks, foster problem-solving, and learn from mistakes without undue blame or scapegoating.

**10. Human-in-the-Loop: Incorporate human input and oversight into all stages of AI decision-making (Accountability, Trust)**

When using an AI tool, there should be a human-in-the-loop, such as a team member or senior executive(s), who takes ultimate responsibility for decision-making informed by AI. This responsibility can shift as an AI system evolves in its lifecycle, with different individuals taking responsibility for their specific duties. The individuals in the loop should also have the authority to intervene at any point when stakeholder feedback or test results indicate that an AI system will result in undesirable outcomes or behaviors. Holding humans accountable for AI decision-making will establish certainty with internal and external stakeholders that humans bear ultimate responsibility for an AI system's output, and in turn, will build trust.

**11. Professional Development: Invest in education and training to upskill and/or reskill employees (Trust, Culture)**

Organizations should support their employees through professional development opportunities that up- or reskill employees with the purpose of working alongside—

rather than resisting, fearing, or resenting—AI technologies. Leadership should demonstrate that professional development is a top priority and encourage and incentivize employees to take advantage of such programs. Further, organizations should implement responsible AI and bias training to increase awareness of their own biases that may emerge when building, testing, using, or monitoring AI systems. These trainings should be specific and applicable to different teams, and organizations should offer rewards for their successful completion.

## **D. ESTABLISH ACCOUNTABILITY AND CLEAR LINES OF RESPONSIBILITY**

Once an organization establishes its responsible AI values and principles, the next step is to design an AI governance framework to operationalize efforts across the enterprise.

EqualAI supports the NIST [Artificial Intelligence Risk Management Framework](#) (NIST AI RMF) to inform and guide an organization in thinking through and laying the foundation for AI governance.<sup>27</sup> Building on the NIST AI RMF, Badge participants pinpointed the need to identify and task key individuals and groups (noted as “tiers” below) with the duties and authority to execute a responsible AI governance framework. The structure below is applicable to any organization regardless of size or industry and can be customized to an organization’s specific needs.

### **► Tier 1: Designate an Authority Who Owns AI Governance**

An organization should start by designating one senior executive who is ultimately responsible for AI governance. This individual could reside in the C-suite (e.g., CIO, CAIO, CLO, COO, or CTO), and if not, they must occupy a position of significant power and influence within an organization. This individual should have the authority to assign or move resources to support a responsible AI mandate.

Efforts to initiate a responsible AI governance framework could also emerge at any level within an organization. For these efforts to be sustained and successful, it is critical to have the support and leadership of the C-suite and board of directors, including this tier 1 authority figure.<sup>28</sup> Importantly, the tier 1 authority figure must embody the leadership skills needed to make difficult decisions that may be unpopular at the time, but necessary to achieve the organization’s responsible AI mandate.

---

<sup>27</sup> National Institute of Standards and Technology, “Artificial Intelligence Risk Management Framework.”

<sup>28</sup> Multiple speakers and participants in the Badge Program touched on the necessity of leadership buy-in from either the C-suite, board of Directors, or both for sustained and effective responsible AI governance efforts.

## ► Tier 2: Institute a Steering or Oversight Committee

To support tier 1 leadership, an organization should appoint and launch a steering committee. This group's mandate is to support the implementation of the responsible AI governance framework and principles by developing an enterprise-wide AI strategy, establishing accountability structures, and managing AI concerns identified by internal and external stakeholders.

The steering committee should have a chairperson who directly reports to and supports the tier 1 authority figure. Members of the steering committee should meet routinely, drive key decisions on AI, and derive from representatives across the enterprise including, but not limited to (and in no particular order): human resources; customer representatives diversity, equity, and inclusion (DEI) leadership; AI development or other technical teams; legal/general counsel; compliance; supply chain; and others who touch on AI production, acquisition, or use, oversee its safety and compliance, or who manage employees or functions where AI is operating to serve or support critical functions.

It is also important to note that steering committee members should not only include individuals with technical skills but, to do this work effectively and bring in the necessary perspectives and expertise, should also include those with social science skills and backgrounds such as human rights or philosophy.

### **Members of the steering committee should be tasked with the following duties:**

- Inform the chairperson and tier 1 authority figure on sectoral and enterprise-wide AI concerns.
- Develop an enterprise-wide strategy to institute the responsible AI governance framework that includes, but is not limited to, initiating regular programming or opportunities to spread awareness (introduction to the framework, workshops to unpack specific elements, etc.)
- Collaborate with the relevant teams to ensure the responsible AI governance framework does not conflict with other internal frameworks (e.g., cybersecurity, privacy), but rather incorporates and/or complements these efforts.
- Document the committee's practical application of the organization's AI principles and framework to help establish precedent with which to understand past decisions and to serve as a guide and a baseline for comparison with future decisions and assessments.

### ► **Tier 3: Appoint Responsible AI Ambassadors**

Responsible AI ambassadors are representatives across the enterprise who, in addition to steering committee members, can understand, explain, and promote the framework. Ambassadors should include representatives from the same teams as steering committee members (see above). This group will be the largest of the three tiers (To clarify: all steering committee members are ambassadors, but not all ambassadors are steering committee members.)

Ambassadors play a key role in building a culture that prioritizes the responsible development and use of AI. They serve as a resource for employees to better understand how to operationalize the framework and execute initiatives stemming from the steering committee, which will help scale efforts for responsible AI governance across the enterprise. To this end, ambassadors are tasked with the following duties:

- Support steering committee members in promoting the responsible AI governance framework and principles within their teams through initiating regular programming or opportunities to share awareness (introduction to the framework, workshops to unpack specific elements, etc.).
- Provide responsible AI training opportunities and professional development programs for their team members.
- Set up a “flag system,” a formal mechanism for employees to notify ambassadors about issues they would like additional training on or about concerns they want to flag to the steering committee.
- Serve as a resource for colleagues who have questions, concerns, or feedback on the framework or principles
- Determine agreed-upon ethical boundaries of what will be built or deployed.

After determining the principles and for each significant development in the responsible AI governance framework and implementing the leadership tiers outlined above, an organization’s leaders should consider hosting a town hall or similar forum to communicate the structure and solicit feedback from employees. Such a forum is important to reassure employees that they are seen and heard on these issues, which will in turn build trust and support positive employee morale.

## E. USE TOOLS TO SUPPORT AI GOVERNANCE

Once values, principles, and a governance framework have been established, organizations should use the following tools, metrics, and monitoring practices to execute and evaluate the success and sustainability of these efforts. The same practices could also be applied to the design and development of AI systems. Throughout the EqualAI Badge Program and Summit, our participants discussed and identified the following methods to best position organizations to effectively develop AI systems and launch responsible AI governance efforts.

### ► Documentation

An organization should standardize how responsible AI management processes are implemented and mandate documentation practices at each stage to ensure consistency and accountability.<sup>29</sup> Documentation encourages knowledge sharing, which empowers employees to understand how processes work and expectations for what a final project or product should look like.<sup>30</sup> The documentation process is especially important with AI systems given that they will touch multiple individuals, entities, and enterprises once deployed.

For powerful AI models, it is crucial that these documentation practices are digestible to both technical and non-technical audiences. Clear and consistent documentation will help inform future users, support effective decision-making, and prepare organizations in heavily regulated industries for audits. **One mechanism to facilitate this effort is through the [EqualAI Algorithmic Impact Assessment \(AIA\) tool](#), which helps organizations check if their AI systems align with the NIST AI RMF.**<sup>31</sup>

### ► Defined Process

An effective corporate process is one where there is a set of defined activities and tasks that, once completed, will accomplish an organizational goal.<sup>32</sup> Establishing clear processes allows organizations to bolster employee esteem and provide a sense of stability and control.<sup>33</sup> This is particularly important in the AI space, where rapidly advancing AI technologies are heavily disrupting various industries.

Organizational processes are only as useful as they are clear to and accepted by key stakeholders. Research indicates that employees will support a manager's decision,

---

<sup>29</sup> Schwartz et al., "Towards a Standard for Identifying and Managing Bias in Artificial Intelligence."

<sup>30</sup> Atlassian, "The Importance of Documentation (Because It's Way More Than a Formality)," <https://www.atlassian.com/work-management/knowledge-sharing/documentation/importance-of-documentation>.

<sup>31</sup> EqualAI, "EqualAI Algorithmic Impact Assessment (AIA)," <https://www.equalai.org/aia/>.

<sup>32</sup> Mary K. Pratt et al., "Business Process," TechTarget, <https://www.techtarget.com/searchcio/definition/business-process>.

<sup>33</sup> Jacob Adler, "Success and Why the Process Matters: Joel Brockner," Wharton Work/Life Integration Project, April 22, 2016, <http://worklife.wharton.upenn.edu/2016/04/success-and-why-the-process-matters-joel-brockner/>.



even if they disagree, if they believe the process the manager used to make the decision was fair. The ever-progressing nature of AI technologies will inevitably yield competing opinions among employees about an organization's decision on AI development and use. As such, establishing processes that employees believe are fair and inclusive will benefit the cohesiveness of the community and reinforce a supportive culture.

Further, in many regulated industries, organizations may be subject to audits and need to demonstrate structure, certainty, and fairness as part of their decision-making processes. While there is no current official industry-wide auditing process for AI systems, organizations should take proactive measures in creating internal processes to best position themselves for future scrutiny. All processes should be documented and digestible for all potential audiences, including technical and non-technical readers.

If an organization is in the business of developing or using AI systems, they must create a process to ensure consistent understanding and application of expectations for the creation and/or adoption of AI systems. This includes prioritizing policies and resources based on the assessed risk level and potential impact of an AI system.

A helpful process to ensure consistent understanding and application of expectations for the creation and/or adoption of AI systems could include the following: identifying the type of AI system in use or expected to be in use, determining intended use-case(s), pinpointing the applicable AI principles, reviewing the ethical dilemmas or concerns employees have flagged, tracking decisions made about the system and why, and analyzing whether the AI system produces outcomes that align with the organization's responsible AI goals. These are just a few examples of steps organizations could string together to create a consistent process for a product's lifecycle. Ultimately, organizations should design processes to fit their needs and goals, and prioritize identifying clearly defined authorities and responsibilities that will help them reach their responsible AI goals.

Organizations should also put forth clear processes and documentation practices that assess the applicability and success of the responsible AI governance framework and principles. Leadership should provide clear expectations for each step of these processes, and identify exactly who is responsible for each step, as well as each overall process.

Further, organizations should establish clear processes for escalation in the event issues arise that require additional oversight or scrutiny.<sup>34</sup> Employees should have the option to remain anonymous if they decide to escalate a concern, and there should be policies in place to prevent retaliation from other employees if an escalation results in a hostile environment.

---

<sup>34</sup> Aaron De Smet, Gerald Lackey, and Leigh M. Weiss, "Untangling Your Organization's Decision Making," McKinsey & Company, June 21, 2017, <https://www.mckinsey.com/capabilities/people-and-organizational-performance/our-insights/untangling-your-organizations-decision-making>.

## ► **Multistakeholder Review**

In designing and building new AI systems, it is important to include multidisciplinary reviews, meaning the broadest variety of perspectives that can help add insight on different potential use cases, biases, and questions that could arise based on views that were not previously explored or integrated. Ideally, these diverse perspectives should be included at every point of the AI development lifecycle to prevent and mitigate potential biases and harms. An effective process includes internal and external stakeholders who stem from a variety of backgrounds, vantage points, and expertise.

Additionally, as an organization develops its responsible AI governance framework, it should proactively seek out input from diverse internal and external stakeholders. These perspectives will help shape a comprehensive and robust framework that best suits the needs of an organization and ensures the broadest variety of perspectives are taken into account. This process is helpful for numerous goals, including best positioning the framework for employee buy-in, as well as accounting for new use cases and end users who were not previously considered. Additionally, after a framework is introduced, there should be routine reviews of how the framework is performing, as detailed further below.

## ► **Metrics, Monitoring, and Reevaluation**

There is not currently a consensus on standard metrics to evaluate or monitor AI systems for risk and harm. A challenge in identifying such metrics includes the fact that a one-size-fits-all approach may be oversimplified, lack critical nuance, and can fail to take into account the differences in affected groups and context.<sup>35</sup> One possible tool organizations can use to help measure and monitor an AI system's risk is the [EqualAI AIA tool](#), which includes a bias and cost-benefit evaluation section that prompts developers to think about potential risks the technology could pose to different populations.<sup>36</sup>

Additionally, there should be a stated cadence of retesting and documenting AI systems, as determined by the product ownership team, which may vary based on the level of risk, use, and other factors. While there is no universally accepted risk-based standard, the NIST AI RMF offers an effective definition of risk: the composite measure of an event's probability of occurring and the magnitude or degree of the consequences of the corresponding event. The impacts, or consequences, of AI systems can be positive, negative, or both and can result in opportunities or threats. When considering the negative impact of a potential event, risk can be a function of 1) the negative impact, or magnitude of harm, that would arise if the circumstance or event occurs; and 2) the likelihood of occurrence.<sup>37</sup> Using this definition, organizations can determine the frequency and rigor of monitoring AI systems as appropriate.

---

<sup>35</sup> National Institute for Standards and Technology, "Artificial Intelligence Risk Management Framework."

<sup>36</sup> EqualAI, "EqualAI Algorithmic Impact Assessment (AIA)."

<sup>37</sup> National Institute for Standards and Technology, "Artificial Intelligence Risk Management Framework."

## **Responsible AI Governance Framework: Metrics for Success**

Identifying metrics to evaluate the success of a governance framework will allow an organization to understand which principles are resonating most with employees and identify areas that need improvement. Such metrics could include:

- The number of ambassadors across the enterprise and diversity of participation (e.g., percentage of business units involved, demographic composition of total governance team, business units and demographic diversity of ambassadors, etc.).
- The number of enrollments and course completions of the Responsible AI Training course established by ambassadors.
- The number of potential risks identified and mitigated.
- The number of alterations to the principles or framework made due to questions posed by ambassadors.
- The number of contacts from other stakeholders (internal/external) to ambassadors/number of AI-related concerns escalated.

Once metrics are established, an organization should establish routine and thorough monitoring practices on its governance framework. This will ensure sustained success and up-to-date principles and governance practices. Further, the effectiveness of the framework should be evaluated at every steering committee meeting. Organizations can use the metrics above to review progress and determine how frequently the committee should meet to ensure their organization's responsible AI goals are met.

# FURTHER DISCUSSION AND OPEN QUESTIONS

---

In addressing organizational values and AI principles and developing a responsible AI governance framework, there were a few areas of deliberation, delineated below, that our participants identified and discussed without reaching a final resolution.

## Deliberative Discussions

### **Centralized vs. Decentralized Steering Committee**

While there was universal consensus that an individual should bear ultimate accountability for responsible AI governance (tier 1 authority figure), there was no consensus on whether the steering committee should have a chairperson who is solely responsible for leading the committee or distributing responsibilities among all steering committee members.

In a centralized steering committee structure, the chairperson leads the steering committee with representatives from across the organization, and engages in periodic communication with key external stakeholders. The chairperson is responsible for and has the authority to oversee the responsible AI governance framework and principles. Other steering committee members will perform their duties in a part-time capacity. While these individuals are held accountable for their duties, the chairperson bears the ultimate responsibility for fulfilling the mandate of the committee.

On the one hand, there is a need for a central steering committee figure to ensure that progress is consistent, efforts are not sidelined due to competing work requirements, and there is a clear line of authority when members do not agree or encounter conflict. On the other hand, if full responsibility falls on a single individual, other members may not be as diligent about fulfilling their duties, which could eventually overwhelm the chairperson.

Alternatively, in a decentralized structure, the steering committee still has a chairperson who works with representatives across the organization and key external stakeholders. However, all members and the chairperson are responsible for fulfilling the steering committee mandate and are held equally accountable for implementing the responsible AI governance framework.

Distributing responsibility could foster more sustainable progress. However, there is a risk of deadlock in progress if conflict arises in the absence of an authority figure for resolution, or that members will inevitably face competing work priorities that cause them to delay or fail to fulfill their responsibilities entirely.

Ultimately, the steering committee structure will depend on a variety of factors, such as the organization's size, industry, and culture.

## **Steering Committee Membership**

There was debate on the a trade-off between involving as many teams and stakeholders as possible and the ability to align and move quickly on decisions on a steering committee.

A large and diverse steering committee could ensure the greatest chance of encapsulating diverse perspectives and reduce the risk of missing crucial viewpoints. This would help mitigate unconscious biases and help members learn from each other about the different factors to consider on any given AI issue. However, more steering committee members will inevitably slow its ability to align on decisions and may prove too cumbersome to act as quickly as needed to keep pace with AI development.

Conversely, a smaller steering committee brings agility as fewer members are needed to align on decision-making. The risk is losing potentially valuable perspectives of stakeholders who are not present that could better inform the committee's decision-making for responsible AI. Other options for achieving a diverse and fresh set of views could be to have rotating seats on the committee (e.g., 18-month terms) for representatives from certain functions (e.g., legal or engineering) and not having tenure/seniority be a barrier to inclusion.

The size of the committee may ultimately depend on the culture of decision-making at an organization. For example, organizations that value debate and collaboration may opt for a larger steering committee to capture more perspectives while organizations with a top-down culture may thrive with a smaller committee where a few key individuals are driving decisions. Additionally, participants deliberated which teams should make up the core of a steering committee but did not reach consensus of what this list looks like. Some teams that were discussed include AI development, legal/general counsel, DEI, human resources, and compliance.

## **Incentives for Employees**

The responsible AI governance framework relies heavily on the involvement of committed and passionate employees. The participants discussed the need for developing incentives to reward employees for successfully fulfilling their roles as ambassadors or steering committee members, but were unable to finalize exactly what these incentives look like. Some ideas that were discussed include an additional annual bonus or adding in performance review metrics used to evaluate for promotions.

# Open Questions

In addition to the discussions above, there are many issues that warrant further scrutiny.

## 1. **Ownership and Advocacy**

The selection process to identify specific steering committee members or ambassadors deserves significant deliberation. Considerations include: Should these individuals be selected based on nominations from their colleagues? Should there be an application process where candidates are interviewed and must provide references? Should they be appointed by the tier 1 authoritative figure?

Additionally, the number of steering committee members and ambassadors remains an open question. Should there be more ambassadors than steering committee members, or vice versa? What are the pros and cons of each option? Further, what should be the minimum time commitment for a steering committee member or ambassador?

Another important issue is how to prevent symbolic or performative participation (i.e., someone is part of the steering committee but fails to fulfill duties) without raising the barrier or deterring employees from getting involved. There is a fine line between a simple pathway to involvement and ensuring that these individuals are committed to doing the work needed to meet the organization's responsible AI goals.

## 2. **Tools, Processes, Metrics, and Monitoring**

Low- and high-risk AI use cases warrant varying levels of scrutiny, but that begs the all-important question of what constitutes a low- or high-risk use case. For example, obvious low-risk and high-risk use cases were brought up, like a TV show-recommending algorithm versus an AI-powered diagnostic healthcare tool, but many use cases are in the gray area and need to be allocated with a risk level for review.

In addition, an organization's risk tolerance must be established. NIST defines this as an organization's readiness to bear risk to achieve its objectives.<sup>38</sup> Risk tolerance would vary based on an organization's industry, sector, and use case, but further discussion on the precise factors that could shape an organization's risk tolerance was not touched upon.

Further, an organization should determine an amendment process for the responsible AI governance framework.

These aforementioned issues are just some of the many pressing issues the EqualAI community plans to scrutinize and unpack at upcoming convenings.

---

<sup>38</sup> National Institute for Standards and Technology, "Artificial Intelligence Risk Management Framework."

# CONCLUSION

---

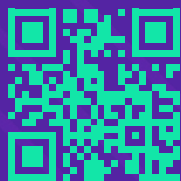
**This white paper attempts to capture a unique and unprecedented event, where a broad cross section of industry leaders congregated at the EqualAI Responsible AI Summit to align on the best practices for responsible AI governance. Such convenings, set against a backdrop of uncertainty about national and international AI standards, are of the utmost importance as organizations cannot afford to wait for certainty to emerge from the legal landscape. Savvy companies, like those listed as co-signatories and participants in this discussion, are taking the lead in setting standards and norms to govern AI responsibly.**

We hope the findings in this white paper are of use to a variety of audiences. For starters, this paper can serve as a beacon for industry leaders who are curious about how other organizations are navigating AI governance intentionally and responsibly. Policymakers can benefit from understanding how responsible industry players are thinking about the risks stemming from AI and how to manage them appropriately. The general public can have a glimpse into how the companies providing their AI-enabled products and services are or should be developing and deploying responsible AI.

We, at EqualAI, feel honored and privileged to be working closely with industry leaders and policymakers who are dedicated to building and using AI responsibly. We look forward to expanding the EqualAI community and building upon the foundational work in this white paper.

# EQUAL™

*EqualAI is a nonprofit focused on reducing bias and other harms in the development and use of Artificial Intelligence (AI) through promotion of responsible AI practices.*



@EqualAI



@ai\_equal