



Six Steps to Bridge the Responsible AI Gap

By Steven Mills, Elias Baltassis, Maximiliano Santinelli, Cathy Carlisi, Sylvain Duranton, and Andrea Gallego

As artificial intelligence assumes a more central role in countless aspects of business and society, so has the need for ensuring its responsible use.

AI has dramatically improved financial performance, employee experience, and product and service quality for millions of customers and citizens, but it has also inflicted harm. AI systems have offered lower credit card limits to women than men despite similar financial profiles. Digital ads have demonstrated racial bias in housing and mortgage offers. Users have tricked chatbots into making offensive and racist comments. Algorithms have produced inaccurate diagnoses and recommendations for cancer treatments.

To counter such AI fails, companies have recognized the need to develop and operate AI systems that **work in the service of good while achieving transformative business impact**—thinking beyond barebones algorithmic fairness and bias in order to identify potential second- and third-order effects on safety, privacy, and society at large. These are all elements of what has become known as Responsible AI.

Companies know they need to develop this capability, and many have already created Responsible AI principles to guide their actions. The big challenge lies in execution. Companies often don't recognize, or know how to bridge, the gulf between principles and tangible actions—what we call crossing the "Responsible AI Gap." To help cross the divide, we have distilled our learnings from engagements with multiple organizations into six basic steps that companies can follow.

The Upside of Responsible AI

Concern is growing both inside and outside boardrooms about the ethical risks associated with AI systems. A survey conducted by the Center for the Governance of AI at the University of Oxford showed that **82% of respondents** believe that AI should be carefully managed. **Two-thirds of internet users** surveyed by the Brookings Institution feel that companies should have an AI code of ethics and review board.

Much of this concern has arisen from failures of AI systems that have received widespread media attention. Executives have begun to understand the risks that poorly designed AI systems can create—from costly litigation to financial losses. The reputational damage and employee disengagement that result from public AI lapses can have far-reaching effects.

But companies should not view Responsible AI simply as a risk-avoidance mechanism. Doing so misses the upside potential that companies can realize by pursuing it. In addition to representing an authentic and ethical “True North” to guide initiatives, Responsible AI can generate financial rewards that justify the investment.

A Stronger Bottom Line. Companies that practice Responsible AI—and let their clients and users know they do so—have the potential to increase market share and long-term profitability. Responsible AI can be used to build high-performing systems with more reliable and explainable outcomes. When based on the authentic and ethical strengths of an organization, these outcomes help build greater trust, improve customer loyalty, and ultimately boost revenues. Major companies such as Salesforce, Microsoft, and Google have publicized the robust steps they have taken to implement Responsible AI. And for good reason: people weigh **ethics three times more heavily than competence** when assessing a company’s trustworthiness, according to Edelman research. Lack of trust carries a heavy financial cost. In the US, BCG research shows that companies lost **one-third of revenue from affected customers** in the year following a data misuse incident.

Brand Differentiation. Increasingly, companies have grown more focused on **staying true to their purpose** and their foundational principles. And customers are increasingly making choices to do business with companies whose demonstrated values are aligned with their own. Companies that deliver what BCG calls **total societal impact (TSI)**—the aggregate of their impact on society—boast higher margins and valuations. Organizations must make sure that their AI initiatives are aligned with what they truly value and the positive impact they seek to make through their purpose. The benefit of focusing strictly on compliance pales in comparison with the value gained from strengthening connections to customers and employees in an increasingly competitive business environment.

Improved Recruiting and Retention. Responsible AI helps attract the elite digital talent that is **critical to the success of firms** worldwide. In the UK, one in six AI workers has quit his or her job rather than having to play a role in the development of potentially harmful products. That’s

more **than three times the rate of the technology sector as a whole**, according to research from Doteveryone. In addition to inspiring the employees who build and deploy AI, implementing AI systems in a responsible manner can also empower workers across the entire organization. For example, Responsible AI can help ensure that AI systems schedule workers in ways that balance employee and company objectives. By building more sustainable schedules, companies will see employee turnover fall, reducing the costs of hiring and training—over \$80 billion annually in the US alone.

Putting Principles into Practice

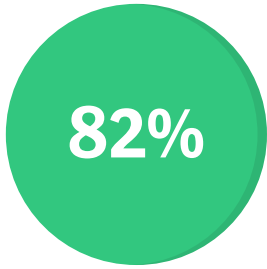
Despite the upsides of pursuing Responsible AI, many companies lack clarity about how to capture these benefits in their day-to-day business. The approach many organizations take has been to create and publicize AI principles. In a comprehensive search of the literature, we found more than 80 sets of such principles. But nearly all lack information about how to make them operational. We consistently observe the gap between principles and practice—the Responsible AI Gap.

Principles without action are hollow. Lack of action not only fails to realize the upside potential of AI but could also be perceived negatively by customers and employees. To cross the Responsible AI Gap, direction is urgently needed. In fact, the practitioners building AI systems are struggling to take tangible action and asking for guidance: **78% percent of tech workers** surveyed by Doteveryone want practical methods and resources to help consider societal impact when building products. (See Exhibit 1.)

Overcoming these challenges requires going far beyond a narrow focus on the algorithms that power AI. Companies must look at every aspect of end-to-end AI systems. They must address front-end practices such as data collection, data processing, and data storage—and pay heed as well to back-end practices, including the business processes in which a system will be used, the decision makers who will determine when and where to implement a system, and the ways information will be presented. They must also ensure that the systems are robust, while keeping top of mind all the potential ways they can fail. They must also address the **large-scale transformation and associated change management** that can generate the greatest impact.

Exhibit 1 - Responsible AI Principles Have Not Altered AI Practices

The public wants Responsible AI ...



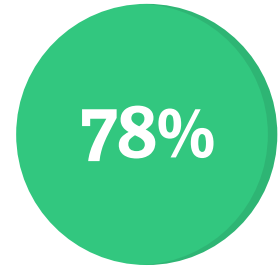
of survey respondents believe AI should be carefully managed

... with robust principles ...



companies have sets of AI principles

... and guidance to put principles into practice



of tech workers want help managing ethical challenges for product development

Sources: Center for the Governance of AI, University of Oxford; Doteveryone; BCG research and analysis.

From our engagements with multiple organizations, we have uncovered six basic steps to make Responsible AI real. While they may appear extensive, leaders should remember that they do not require a massive investment to get started. Each step can begin small and evolve and expand over time as an initiative matures. The important thing is that organizations should make progress across each of these steps. (See Exhibit 2.)

1. Empower Responsible AI leadership. An internal champion such as a chief AI ethics officer should be appointed to sit at the helm of the Responsible AI initiative. That leader convenes stakeholders, identifies champions across the organization, and establishes principles and policies that guide the creation of AI systems. But leadership with ultimate decision-making responsibility is not enough. No single person has all the answers to these complex issues. Organizational ownership that incorporates a diverse set of perspectives must be in place to deliver meaningful impact.

A powerful approach to ensuring diverse perspectives is a multidisciplinary Responsible AI committee that helps steer the overall program and resolve complex ethical issues such as bias and unintended consequences. The committee should include representation from a diversity of business functions (e.g., BUs, public relations, legal, compliance, AI team), regions, and backgrounds. A recent

BCG study suggests that increasing the diversity of leadership teams leads to more and better innovation and improved financial performance. The same is true for Responsible AI. Navigating the complex issues that will inevitably arise as companies deploy AI systems requires the same type of diverse leadership.

2. Develop principles, policies, and training. Although principles are not enough to achieve Responsible AI, they are critically important, since they serve as the basis for the broader program that follows. Responsible AI principles should flow directly from the company's overall purpose and values to provide clear links to corporate culture and commitment. (See Exhibit 3.) Time must be invested to develop, socialize, and disseminate Responsible AI principles. The process of soliciting broad feedback from across the organization has the added benefit of identifying employee concerns and areas of particularly high risk. Ensuring that principles are communicated broadly provides employees with the context for initiatives that will follow.

A press release or companywide email is not enough to make principles real. Principles must be broken down into specific and actionable policies and standards around which teams can execute. Without these details, companies may fail to translate principles into tangible actions.

Exhibit 2 - Six Steps to Make Responsible AI Real



Empower Responsible AI leadership

Appoint a leader and a diverse team to design and lead the Responsible AI program and drive initiatives.



Develop principles, policies, and training

Build, communicate, and disseminate Responsible AI principles, policies, and training to all members of the AI team, including leaders.



Establish human + AI governance

Establish roles and responsibilities, a mechanism for review and adherence, escalation paths to raise concerns, and accountability for outcomes.



Conduct Responsible AI reviews

Build or adopt a tool for conducting end-to-end use case reviews and ensure they are conducted at scale.



Integrate tools and methods

Evolve standard data, technology, and model building to include Responsible AI considerations.

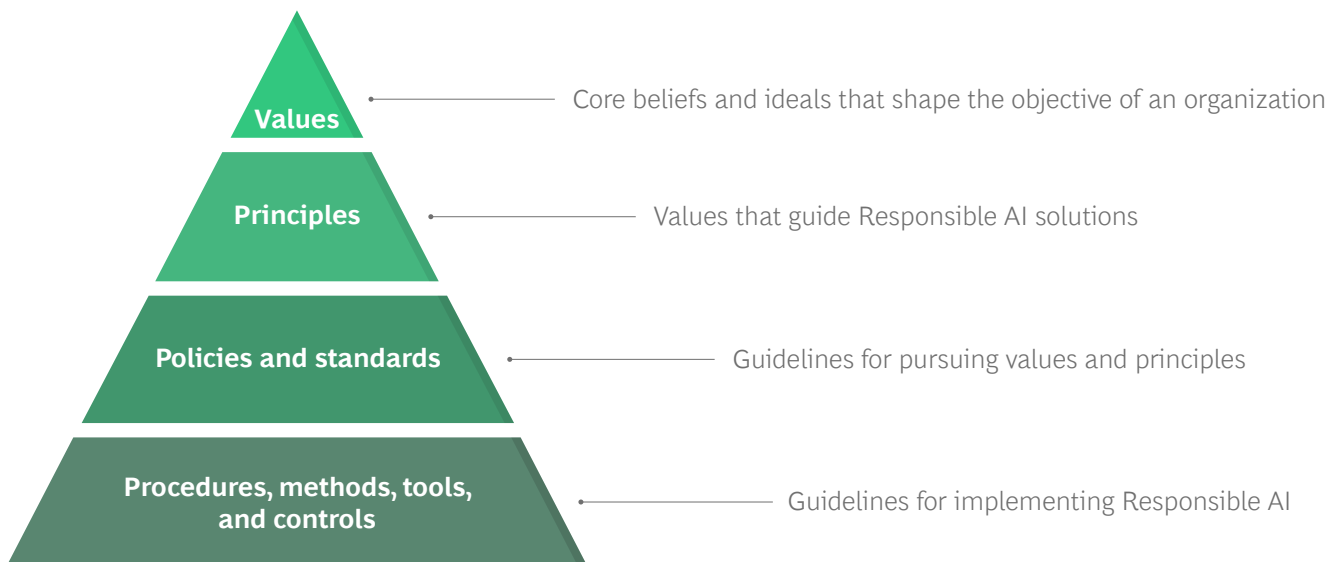


Build and test a response plan

Create the roles and responsibilities, processes, and procedures to respond when a Responsible AI lapse occurs, as well as to periodically test and refine.

Source: BCG RAI.

Exhibit 3 - A Framework for Defining Responsible AI Initiatives



Source: BCG RAI.

Ethical practices require systematic communications and training to educate teams about Responsible AI and a company's specific approach. Training needs to go beyond AI system developers to span all levels and areas of the business—from the C-suite to the end users of AI systems. Ultimately, Responsible AI is a shared commitment. Everyone has a role to play. Organizations need to foster an open “see something, say something” culture so that issues are identified and raised and that honest dialogue occurs around these complex and often sensitive matters.

3. Establish human + AI governance. Beyond executive leadership and a broadly understood ethical framework, roles, responsibilities, and procedures are also necessary to ensure that organizations embed Responsible AI into the products and services they develop. Effective governance involves bridging the gap between the teams building AI products and the leaders and governance committee providing oversight, so that high-level principles and policies can be applied in practice.

Responsible AI governance can take a variety of forms. Elements include defined escalation paths when risks emerge at a particular project stage, standardized code reviews, ombudspersons charged with assessing individual concerns, and continuous improvement to strengthen capabilities and confront new challenges.

4. Conduct Responsible AI reviews. For Responsible AI to have an impact, the approach must be integrated into the full value chain. Effective integration hinges on regularly assessing the risks and biases associated with the outcomes of each use case. Using a structured assessment tool helps identify and mitigate risks throughout the project life cycle, from prototype to deployment and use at scale. By assessing development and deployment at every step of the journey, AI practitioners can identify risks early and flag them for input from managers, experts, and the Responsible AI governance committee. These reviews should not be limited to the algorithms but be a comprehensive assessment of the complete, end-to-end AI system, from data collection to users acting on the recommendations of systems.

For example, a company might develop a recruiting model that prioritizes candidates for interviews based on their likelihood of receiving job offers. However, after using the assessment tool, the organization might realize that training the model on historical data from job applications is biased due to an underrepresentation of minority groups. Through a combination of data preparation, model tuning, and training for recruiters, the AI system could help increase the diversity of the candidate pool. Or consider an assessment of an e-commerce personalization engine that

uses past purchase behavior and credit history to recommend products on a luxury retail website. The assessment tool could help the team identify that the engine systematically promotes products to lower-income individuals. A broader discussion could be triggered about the potential unintended consequences of a system that encourages purchases that might worsen the financial situation of some customers.

5. Integrate tools and methods. For Responsible AI principles and policies to have an impact, AI system developers must be armed with tools and methods that support them. For example, it is easy for executive leaders to require teams to review data for bias, but conducting those reviews can be time-consuming and complex. Providing tools that simplify workflows while operationalizing Responsible AI policies ensures compliance and avoids resistance from teams that may already be overloaded or operating under tight deadlines.

Toolkits comprising tutorials, modular code samples, and standardized approaches for addressing common issues such as data bias and biased outcomes are important resources. These learning resources should be made available to everyone involved with AI projects so that they can be applied to different contexts and guide individuals designing AI systems. Companies cannot require technical teams to address nuanced ethical issues without providing them with the tools and training necessary to do so.

Creating these resources may sound like a substantial undertaking. While that may have been true a few years ago, a variety of commercial and open-source tools and tutorials are now available. Instead of building their own resources, companies can begin by curating a set of resources that are most applicable to the AI systems they develop. Over time, resources can be customized to a company's specific needs in ways that limit a large up-front investment.

6. Build and test a response plan. Preparation is critical to making Responsible AI operational. While every effort should be taken to avoid a lapse, companies also need to adopt the mindset that mistakes will happen. A response plan must be put in place to mitigate adverse impacts to customers and the company if an AI-related lapse occurs. The plan details the steps that should be taken to prevent further harm, correct technical issues, and communicate to customers and employees what happened and what is being done. The plan should also designate the individuals responsible for each step, so as to avoid confusion and ensure seamless execution.

Procedures need to be developed, validated, tested, and refined to ensure that if an AI system fails, harmful conse-

**Responsible AI principles should
flow directly from the company's
overall purpose and values**



quences are minimized to the greatest extent possible. A tabletop exercise that simulates an AI lapse is one of the best tools companies can use to pressure-test their response plan and practice its execution. This immersive experience allows executives to understand how prepared the organization is and where gaps exist. It's an approach that has proven effective for [cybersecurity](#) and can be equally valuable for Responsible AI.

An Opportunity for Growth

While our approach may sound demanding, we firmly believe that delivering AI responsibly is achievable for any organization. Each step does not in itself require a massive initiative or investment. Companies can start with a small effort that builds over time.

The specific approach to the six steps we have described will differ depending on each organization's individual context, including its business challenges, organizational culture, values, and legal environment. Nevertheless, the fundamentals remain the same. Ethical leadership is critical, as is the establishment of broad-based support for the required internal change.

Fortunately, putting Responsible AI into practice does not mean missing out on the business value AI can generate. This is not an "either/or" issue, but rather a "both/and" opportunity in which Responsible AI can be realized while still meeting—and exceeding—business objectives. But for AI to achieve a meaningful and transformational impact on the business, it must be grounded in an organization's distinctive purpose. Only in that way can an organization build the transparency and trust that binds company and customer, manager and team, and citizen and society.

About the Authors

Steven Mills is a partner and associate director in the Washington, DC, office of Boston Consulting Group and the Chief AI Ethics Officer and global lead for the Artificial Intelligence topic in Public Sector. In addition, he is a senior member of BCG GAMMA, the firm's global team of world-class data scientists. You may contact him by email at mills.steven@bcgfed.com.

Elias Baltassis is a partner and director in the firm's Paris office and a senior member of BCG GAMMA, where he leads the topics of Data Strategy & Governance and Analytics Organization. You may contact him by email at baltassis.elias@bcg.com.

Maximiliano Santinelli is an associate director in data science in BCG's Boston office. You may contact him by email at santinelli.maximiliano@bcg.com.

Cathy Carlisi is a managing director and president for the Americas of BrightHouse, BCG's global creative consultancy that pioneered the practice of helping organizations uncover their timeless purpose to grow their people, profits, and social impact. She is based in BCG's Atlanta office. You may contact her by email at ccarlisi@thinkbrighthouse.com.

Sylvain Duranton is a managing director and senior partner in the firm's Paris office and the global leader of BCG GAMMA. You may contact him by email at duranton.sylvain@bcg.com.

Andrea Gallego is a partner and managing director in BCG's Boston office, the chief technology officer of BCG GAMMA, and a founder of Source.AI, an AI and advanced analytics platform. You may contact her by email at gallego.andrea@bcg.com.

Boston Consulting Group partners with leaders in business and society to tackle their most important challenges and capture their greatest opportunities. BCG was the pioneer in business strategy when it was founded in 1963. Today, we help clients with total transformation—inspiring complex change, enabling organizations to grow, building competitive advantage, and driving bottom-line impact.

To succeed, organizations must blend digital and human capabilities. Our diverse, global teams bring deep industry and functional expertise and a range of perspectives to spark change. BCG delivers solutions through leading-edge management consulting along with technology and design, corporate and digital ventures—and business purpose. We work in a uniquely collaborative model across the firm and throughout all levels of the client organization, generating results that allow our clients to thrive.

© Boston Consulting Group 2020. All rights reserved. 9/20

For information or permission to reprint, please contact BCG at permissions@bcg.com. To find the latest BCG content and register to receive e-alerts on this topic or others, please visit bcg.com. Follow Boston Consulting Group on Facebook and Twitter.