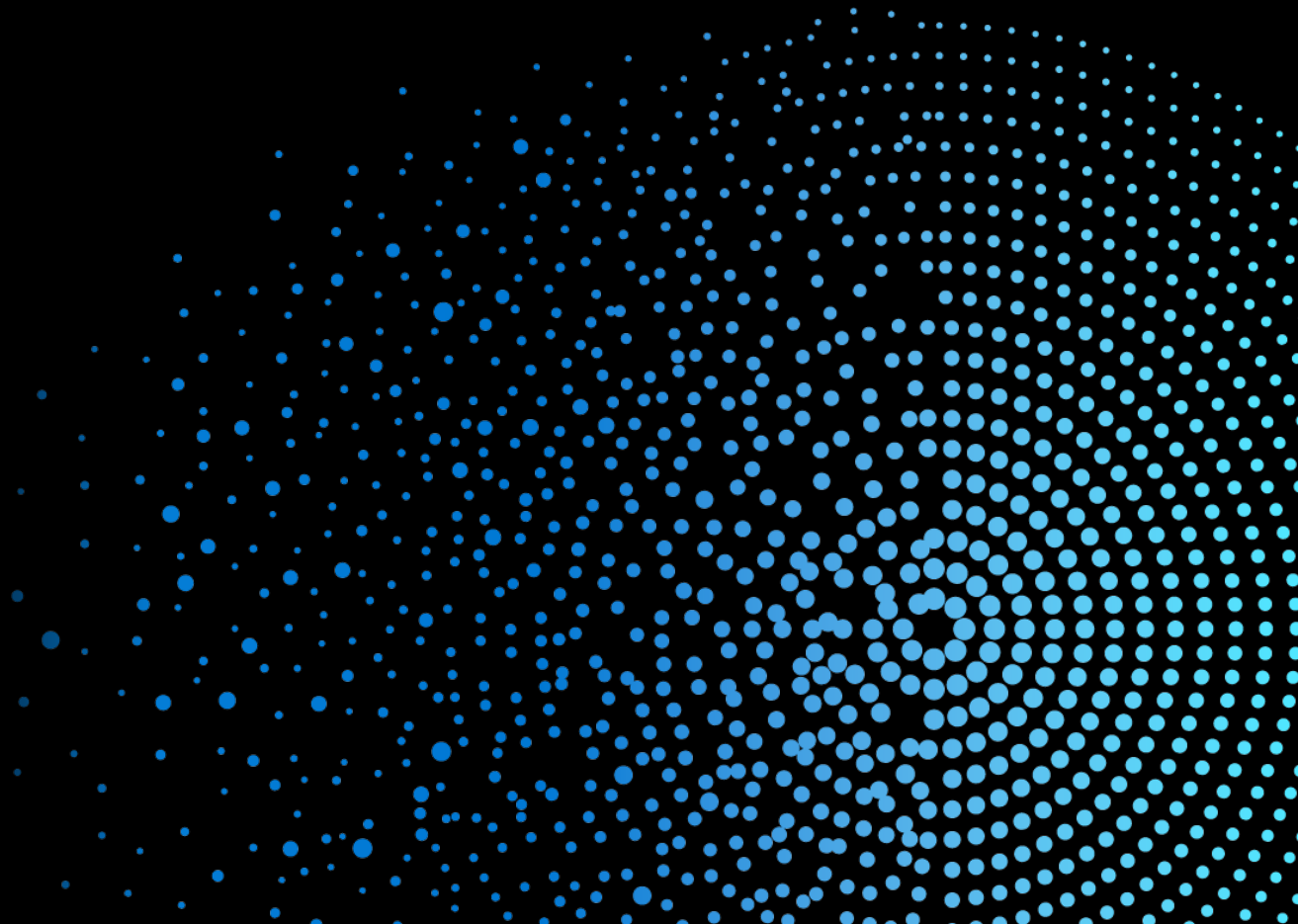


Ten Guidelines for Product Leaders to Implement AI Responsibly



Ten Guidelines for Product Leaders to Implement AI Responsibly

As AI gets ever more deeply embedded in society and people's everyday lives, business leaders, data scientists, and engineers have a responsibility to guide AI product development toward responsible outcomes. In this paper, we provide some considerations and prompts to help product leaders, including lead developers, designers, and system architects, chart such a course.

Product leaders typically oversee the development of one or more AI products. Their teams have simultaneous responsibility to advance an organization's values and business objectives while guiding product development toward responsible outcomes. Their place at the center of innovation and product development positions these leaders as essential actors in operationalizing responsible AI. The range of AI products they oversee is broad, with the underlying algorithms ranging from novel, complex models to pre-trained models built by others. The guidelines in this paper provide an approach to help product leaders responsibly navigate AI product development. While elements of these guidelines apply to a variety of circumstances, we have created them specifically to serve teams that are building AI models, not for those who are procuring finished products or pre-trained models.

Organizational leaders focused on product development and delivery should remember to take sufficient measures to scale responsible AI across their organization. Ultimately, responsible AI is about driving a cultural shift within an organization. Truly operationalizing responsible AI requires broad changes across leadership, governance, processes, and talent. Given their proximity to product teams, product leaders are key actors in catalyzing the broader changes, but it requires them to focus across the vectors. While managing all these changes is beyond the scope of this report, additional resources are available for business leaders navigating this journey.¹

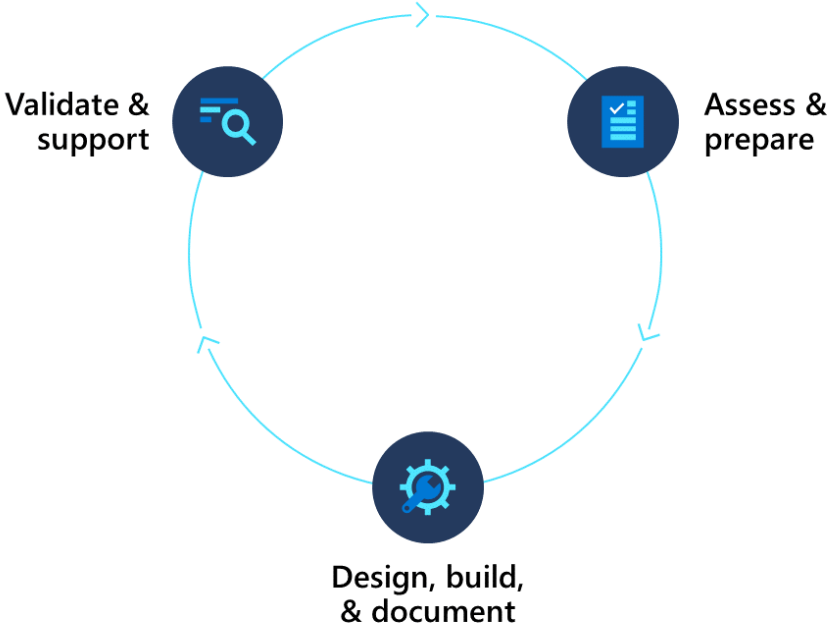
¹ For example, see Microsoft's [Discover ways to foster an AI-ready culture in your business](#) and "[This is why we need to talk about Responsible AI](#)," World Economic Forum (2020).

Responsible AI Guidelines for Product Leaders



The ten guidelines outlined below focus on product development and delivery processes to provide a formalized way to stimulate the challenging conversations and critical thinking necessary to anticipate and mitigate the risks of AI systems. Since every use case brings a unique context and set of challenges, the guidelines are not intended to be used as a checklist or to prescribe specific design choices. Furthermore, organizations must view product development in the context of a rapidly changing regulatory environment. As such, these guidelines should not be viewed as a tool for achieving regulatory or legal compliance. Teams should always work with relevant internal departments to ensure that the AI products they create adhere to all applicable laws and regulations within the jurisdictions where the product will be developed, used, or marketed.

We believe that these guidelines will help product leaders create deeper synergy between performance, organizational goals, and values. Throughout the product development lifecycle, leaders should keep in mind their organization’s values and AI ethical principles, to the extent their organization has formalized them. Inevitably, issues will arise that require product teams to make judgment calls. In these situations, product leaders should use values and principles as a north star to guide their thinking. They should also seek diverse perspectives from within and outside their organization to provide them with input from all relevant stakeholders.

The guidelines have been organized according to the key phases of the product development lifecycle while recognizing that AI product development often cycles through these phases iteratively:



Ten Guidelines for Product Leaders to Implement AI Responsibly

ACTION	GUIDELINES
 Assess & prepare	<ol style="list-style-type: none">1 Assess merit of developing the product considering organizational values and business objectives2 Assemble team reflecting diverse perspectives and with clearly defined roles and responsibilities3 Assess potential product impact by including input from domain experts and potentially impacted groups
	<ol style="list-style-type: none">4 Evaluate data and system outcomes to minimize risk of fairness harms5 Design AI product to mitigate potential negative impact on society and environment
	<ol style="list-style-type: none">6 Incorporate features to enable human control7 Take measures to safeguard data and AI products8 Document throughout the development lifecycle to enable transparency
 Design, build & document	<ol style="list-style-type: none">9 Validate product performance and test for unplanned failures as well as foreseeable misuse unique to AI products10 Communicate design choices, performance, limitations, and safety risks to end user
	<ol style="list-style-type: none">9 Validate product performance and test for unplanned failures as well as foreseeable misuse unique to AI products10 Communicate design choices, performance, limitations, and safety risks to end user
 Validate & support	

Applying the Guidelines

Along with each guideline, we have provided a series of questions to help product leaders engage their teams on responsible AI. The questions are intended to spur critical thinking and help teams proactively surface risks. Certain questions will apply to some products more than others, but they are designed to be broadly applicable. The questions are not all-inclusive, but they do reflect best practices drawn from Microsoft and BCG's experience developing and delivering responsible AI solutions. Several illustrative use cases are included to demonstrate how product teams can use these questions to deliver responsible solutions.

Across all phases of product development (e.g., assess and prepare), product leaders should regularly convene their product team, discuss relevant questions, and review prior answers. If the team cannot answer one of the questions with sufficient clarity, the team should invest time in developing a specific action plan. Product teams should view these best practices as a starting point for discussion and planning.



Assess & prepare

1

Assess merit of developing the product considering organizational values and business objectives

- What are the primary use cases and benefits of the proposed product? Which uses are explicitly out of scope?
 - What is the desired business outcome for this product? How will the business impact be measured?
 - How might the operation of the product and use of its outputs for business decisions encroach on core organizational values?
-

2

Assemble team reflecting **diverse perspectives** and with clearly **defined roles and responsibilities**

- Do we have a diverse (e.g., gender, age, ethnicity), multi-disciplinary team with a range of functional expertise?
 - What perspectives or expertise are missing and how can we introduce them, including sources outside the team or organization?
 - Is the team structured so that domain experts can impact relevant design choices?
-

3

Assess potential product impact by including input from **domain experts and potentially impacted groups**

- What are the foreseeable modes of failure for this product? What edge scenarios could lead to failure and harm?
 - What are the societal and environmental implications of foreseeable product failure, misuse, or malicious attack?²
 - What are the product's potential unplanned uses?
 - What external SMEs or groups could provide input informing design choices that would reduce the risk of negative societal impact and harm to individuals directly or indirectly affected by the AI product?
-

² Failure is defined as an inability to perform its expected function either through design flaw or subversion.

Assess & prepare: Automating loan approval decisions

A financial institution would like to automate its loan application approval process with the goal of increasing efficiency and reaching new customer bases. The product leader immediately recognizes that, because of the AI product's potential to impact individuals' economic wellbeing and quality of life, they must proceed with care.

Assess merit of developing the product considering organizational values and business objectives

Before assembling the team, the product leader emails a data scientist and a user researcher with whom they've worked with in the past to talk through potential approaches. After an hour of whiteboarding, they conclude with the following assessment:

The use case is designed to predict an applicant's earning potential over the life of the loan under review, and thus the likelihood of repayment.

The business impact should be measured with three KPIs: default rate against a historical baseline, increase in overall loan volume, and speed of approval/rejection decision.

There may be race, gender, or other biases reflected in the historic data. Exploratory data analysis will need to include steps to bias assess the dataset composition so that appropriate mitigations can be implemented.

There is a significant risk that proxy variables for race, gender, and other protected categories could reinforce biases and impact outcomes, violating company values around fairness and the law. The product team will have to proactively engage on these topics.

Assemble team reflecting diverse perspectives and with clearly defined roles and responsibilities

The product leader begins to build the team based on the initial assessment. The team should include:

Data scientists with experience applying fairness tools to ML models.

A lawyer with expertise in the Equal Credit Opportunity Act to make sure the team fully understands the regulatory environment.

Two loan officers with extensive experience helping a variety of customers navigate the application process. This will enable the team to design the product to augment the decision making—capabilities of loan officers and integrate their feedback in real-time.

A product leader who knows that, since the system will deal with sensitive personally identifiable information (PII), system security and privacy are critical. The leader must gain approval to bring on an external consultant with cybersecurity and AI expertise to provide guidance on optimal system architecture, data storage, and differential privacy across the build.

A user researcher to ensure customer needs are at the forefront, from idea inception to final product, so that customer understanding and satisfaction can be reconciled with business goals.

A designer to improve product usability and accessibility across a diverse user base.

A diverse team reflecting a variety of backgrounds and lived experiences.

Assess potential product impact by including input from domain experts and potentially impacted groups

During a preliminary discussion, the loan officers share their experiences engaging with the financial institution's current customer base. As they detail interactions with various customers over the years, the product leader realizes that, hopefully, the AI may engage new customer demographics who may have needs and expectations that differ from those captured in historical data.

The team engages an economist with expertise in banking relationships within the demographic groups to whom the bank might expand its activities.

The team also works with a vendor to deploy a survey for potential borrowers in new customer communities to better understand how greater access to credit might impact earning potential—in ways both consistent and inconsistent with communities overrepresented in the historical data—and thus the likelihood of repayment.

The loan officers note that even with a new customer base, one thing is likely to remain constant: Having a loan application declined is an unpleasant and potentially painful experience, one that can be delivered more respectfully by a skilled and experienced professional. Based on this insight, the product team decides that:

All rejection decisions will be communicated to applicants by a loan officer.



Design, build, & document

4

Evaluate data and system outcomes to **minimize the risk of fairness harms**

- What fairness metrics (e.g., statistical parity, equalized odds), tests, and shipping criteria will we use? How will the product team validate that the training data, including data collected via APIs, captures the different groups and types of people likely to be impacted by the system's output?
- How will the product team measure whether the AI product's outcomes are consistent with the chosen objective (i.e., avoid target leakage), fairness metrics, tests, and shipping criteria across a wide variety of potentially impacted groups or intersections of groups?
- How will the product team ensure continued adherence to fairness metrics, tests, and criteria post-deployment?

5

Design AI product to **mitigate the potential negative impact on society and the environment**

- If negative impact (e.g., from system failure, unplanned use, abuse, attack, or simply side effect of normal use) is possible, what design processes (e.g., human-centered-design) and choices can reduce, mitigate, or control them?
 - What design choices will help minimize the adverse environmental impact of the product outputs and related decisions? What design choices are critical to ensuring proper use, legitimate and transparent data collection, and respect for user privacy?
-

6 Incorporate features to enable **human control**

- How is the product team designing the product to empower humans by augmenting their decision-making, streamlining tasks, or otherwise making them more effective? Which decisions or functions require human oversight as a critical component of the AI product?
- What mechanisms (e.g., interpretability) will support end user comprehension of the system to enable continuous audit, monitoring, and human intervention?
- What product features allow users to customize AI performance?
- What channels will the product utilize to collect live feedback?
- What product features will ensure inclusive experiences for people with disabilities?

7 Take measures to **safeguard data and AI products**

- How will the product team prioritize data privacy across product design with respect to cloud infrastructure, encryption, anonymization, and access control?
- How will the product team make sure the product doesn't inadvertently disclose sensitive or private information during use (e.g., indirectly inferring user locations or behavior)?
- What methods will the team use to identify and address security vulnerabilities, including those such as data poisoning that are unique to AI products?

8 Document throughout the development lifecycle to **enable transparency**

- What standards and processes are in place to ensure the entire product team consistently documents design and development choices, rationales, and assumptions?
 - How can the product team best keep track of data sources and their authorized uses?
 - What types of models, tools, or techniques will we use to document product behavior?
-

Design, build, & document: Demand forecasting for a fashion retailer

A fashion retail chain hopes to transform its in-store inventory management with AI. Using historical sales data, the company wants to optimize the amount of inventory held in stores so that it can maximize sales per square foot. The product leader has clarified the business objective and assembled a team. The team is now shifting its focus to the design, build, and document stage of work.

Evaluate data and system outcomes to minimize the risk of fairness harms

The product leader convenes the product team for a discussion around fairness. A poorly designed product could lead to service discrepancies across different demographics within the customer base, which would violate the company's values. In brainstorming potential challenges, a data scientist on the team notes that customer feedback for the retail chain varies dramatically across neighborhoods in large metropolitan areas, with stockouts more common in some locations than others. Pivoting from that comment, the team aligns on the following approach:

The team decides to check for parity of service level based on stockouts reported in its historical data. Cross-referencing the results with census data will allow the team to assess whether stockouts reported are correlated with certain neighborhoods and, perhaps, demographic groups.

To prevent unacceptably low service levels in specific locations, the product team establishes a minimum inventory level for each SKU at each location to ensure minimum service levels that will avoid the possibility that stockouts would impact certain demographic groups more than others.

Design AI product to mitigate the potential negative impact on society and the environment

The team's sustainability expert asks the group if system usage might have some second-order environmental effects. Optimizing inventory levels could maximize profit at the store level, but smaller and more frequent inventory replenishment would rely on greater air and ground cargo traffic. At the country level, the environmental consequences of that cargo traffic could be significant. Furthermore, since significant amounts of excess inventory would need to be returned by stores, optimized inventory levels could create additional cargo traffic.

Using internal logistics data, the product team builds a feature that highlights tradeoffs between inventory levels and transportation emissions at the store and region level.

Incorporate features to enable human control

A specialist in human-centered design observes that the retailer often learns about new trends from its frontline workers, particularly store managers. The historical data has limited predictive power for trend spotting. By enabling human control over the inventory system and augmenting store managers' decision-making abilities, stores could make on-the-fly adjustments to match changing consumer preferences. Based on further exploration, the team decides to:

Design a feedback mechanism through which store managers can indicate emergent trends and newly popular products at a specific location, thus allowing for the pooling of insights across the country to help spot trends and adjust inventory levels accordingly.



Validate & support

9

Validate product performance

and test for **unplanned failures as well as foreseeable misuse** unique to AI products

- How will the product team validate the AI product's performance against agreed-upon business KPIs and metrics, tests, and criteria?
- How will the product team validate the AI product's performance against technical standards and benchmarks?
- What are the target environment and conditions under which this product can be expected to function properly and safely?
- How will the system be tested and evaluated for safe and effective operations (e.g., graceful failure) in both business-as-usual and edge-case scenarios?
- What are the mechanisms for continuously monitoring business, technical, and fairness performance, as well as for modeling drift post-production?

10

Communicate design choices, performance, limitations, and safety risks to end user

- What information and instructions should the product team provide to the end user(s) to enable safe and reliable use?
 - How will the product team make sure end users understand the primary use case, underlying assumptions, and limitations of the product?
 - How will the outputs of the system be communicated in a way that helps end users understand how the system works?
-

Validate & support: Predictive lead times for a manufacturer

An industrial goods manufacturer is experiencing repeated delays in the delivery of parts purchased from suppliers. This has caused disruptions to the manufacturing schedule and, ultimately, late deliveries to customers—and the potential to damage key relationships. The manufacturer has decided to develop an AI product to estimate lead times for supplier-procured components based on historical data. Providing purchasing managers with an early warning of potential delays will enable them to proactively engage with suppliers and adjust the manufacturing schedule accordingly to avoid missed delivery deadlines.

Validate product performance and test for unplanned failures as well as foreseeable misuse unique to AI products

This AI product would change the way the company engages with suppliers, customers, and workers. Although the product is expected to have positive impacts across operations and customer relations, erroneous lead time estimates could also create more work for purchasing managers and further damage supplier relationships and negatively impact the bottom line. This could have significant and far-reaching consequences for the business. Furthermore, because assembly is a labor-intensive process requiring specialized skills and safety certifications at different stages, ad hoc adjustments to the manufacturing schedule could place factory employees at risk.

The product leader gathered the product team to pose a series of questions to reach alignment on the approach to validating product performance and robustness to unplanned failures. The meeting concluded with an agreement on a path forward that included:

- Testing the system's outputs against historical supplier-promise dates to determine the model's ability to flag potentially delayed shipments before it's too late.

- Inputting a wide range of operational scenarios, including varied suppliers and component types, as well as edge scenarios (e.g., parts and suppliers not found in the historical data).

- For each scenario tested that requires adjustments to the manufacturing schedule, developing associated work schedules to be validated by factory floor leadership for feasibility and safety.

During the discussion of operational scenarios, a team data scientist noted that factors related to the ongoing COVID-19 pandemic, such as limited trucking capacity and economic shutdowns in certain states, would not be captured in historical data—but could impact lead times. To capture these insights, the product team integrated COVID caseloads in geographies proximate to suppliers as a way to capture the pandemic's potential impact on the manufacturer. In conversations with factory floor leadership, the product team learned how scheduling had recently changed to minimize risks of COVID-19 exposure, and that any changes to the manufacturing schedule would have to be consistent with the new scheduling policies.

Communicate design choices, performance, limitations, and safety risks to end user

The team's UX lead pushes the product team to consider how best to augment the purchasing managers' current decision-making process. They start brainstorming ideas. Others join the discussion. The team aligns on the following steps:

- Calculate confidence intervals alongside estimated lead times to enable purchasing managers to responsibly leverage system outputs.

- Design the dashboard to sort components for a certain product by predicted delay, focusing the end-users' attention on prioritizing key products, and engaging problem suppliers early.

- Make sure the system never automatically updates customers on delivery dates, a task that program managers themselves should continue to perform.

- Build an additional supplier-level, as opposed to component-level, dashboard to equip the company for strategic engagement with suppliers that consistently struggle to deliver parts on time.

- Output a draft work schedule based on proposed changes that will support purchasing manager decision making regarding adjustments to the manufacturing schedule. The factory floor leadership will have to sign off on the revised work schedule to ensure adherence to worker-safety standards.

The team, having earlier identified and engaged with the product's end users, conducts several meetings and trainings to ensure that these features will be leveraged effectively. It also designs a modular training on tool use to be integrated into future onboardings programs for purchasing managers.

Conclusion

There are many building blocks to developing AI responsibly. Broadly speaking, securing the potential of AI in ways that benefit everyone will depend on successful collaboration between governments, industry, and civil society. Within organizations, however, leaders play a critical role in preserving an AI perspective that is both people-centric and promotes positive social impact. Product leaders can be change agents in their organizations' cultural transformations. They can also play an important role in catalyzing and facilitating the complex and sometimes uncomfortable conversations that are necessary to responsibly develop products.

Under such circumstances, team members should be made to feel comfortable raising questions around sensitive issues and identifying gaps in expertise and experience. The guidelines in this paper provide product leaders with prompts to guide these critical conversations during the product-development lifecycle. The cross-pollination of ideas among people with diverse backgrounds and expertise can deliver results that bestow the greatest benefit to society while providing transformative business value.

Reference

¹For example, see Microsoft's [Discover ways to foster an AI-ready culture in your business](#) and "[This is why we need to talk about Responsible AI](#)," World Economic Forum (2020).

²This definition has already been included as a footnote.

© 2021 Microsoft Corporation. All rights reserved. This document is provided "as is." Information and views expressed in this document, including URL and other internet website references, may change without notice. You bear the risk of using it. This document does not provide you with any legal rights to any intellectual property in any Microsoft product. You may copy and use this document for your internal, reference purposes.