

# A practical guide to Responsible Artificial Intelligence (AI)

# Foreword

Everyone's talking about responsible AI. To turn the talk into action, organisations need to make sure that their use of AI fulfills a number of criteria. First, that it's ethically sound and complies with regulations in all respects; second, that it's underpinned by a robust foundation of end-to-end governance; and third, that it's supported by strong performance pillars addressing bias and fairness, interpretability and explainability, and robustness and security.



From automation to augmentation and beyond, artificial intelligence (AI) is already changing how business gets done. Companies are using AI to automate tasks that humans used to do, such as fraud detection or vetting resumés and loan applications, thereby freeing those people up for higher-level work. Doctors are using AI to diagnose some health conditions faster and more accurately. Chatbots are being used in place of customer service representatives to help customers address simple questions.

Through all these roles and more, AI opens up virtually limitless potential to benefit the whole of society. Businesses in every sector are eager to claim their piece of the potential AI windfall, which [PwC research](#) estimates could translate to an infusion of US\$15.7 trillion into the global economy by 2030<sup>1</sup>.

PwC's recently released Responsible AI Diagnostic surveyed around 250 senior business executives from May to June 2019. It found that the level of understanding and application of responsible and ethical AI practices among respondents significantly varied across organisations, and in most cases was immature. The findings also highlighted challenges around access to the skills needed to adopt responsible AI practices.

Of businesses currently using or trialling AI solutions:

**25%**

have not considered AI as part of their corporate strategy

Only **38%**


think it is aligned with their organisational values

Only **25%**

definitely prioritise a consideration of the ethical implications of an AI solution before investing in it

<sup>1</sup> PwC's Global Artificial Intelligence Study: Sizing the Prize



A close-up photograph of a woman with dark, curly hair and glasses, looking intently at a tablet device. She is wearing a colorful, patterned top. The background is a blurred office environment with soft lighting.

# AI concerns differ between consumers and business leaders

Amid this promise, the rapid pace and significant scale of change resulting from ever smarter AI systems and increasingly pervasive human/machine interactions are also giving rise to markedly differing concerns among business leaders and consumers.

Consumers want the convenience of services tailored to their needs, together with the peace of mind knowing that companies are not unknowingly biased against them—and that their government will protect them with laws regulating how their data can be used. Businesses, meanwhile, are in many cases still exploring the opportunities AI presents and, at the same time, educating themselves about the possible risks.

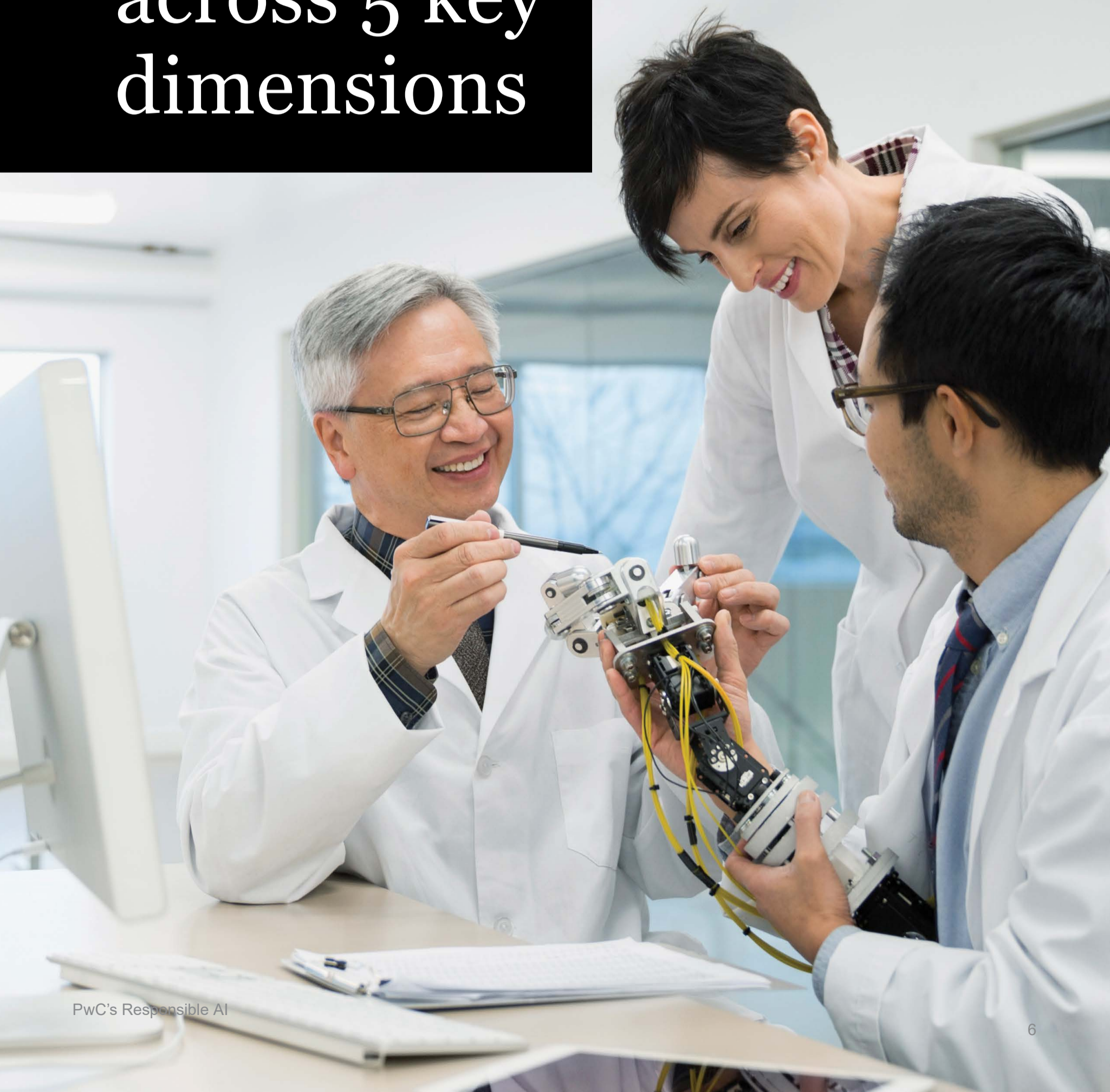
**84% of CEOs agree that AI-based decisions need to be explainable in order to be trusted.**

PwC's most recent [Global CEO Survey](#) found that the risks as well as opportunities around AI are a key focus for top executives. Eighty-five percent of CEOs agree that AI will significantly change the way they do business in the next five years, and close to two-thirds consider AI to be bigger than the internet revolution. But on questions around how much AI can be trusted, opinions are less clear-cut. Over three-quarters of CEOs think AI is “good for society,” but an even higher proportion—84%—agree that AI-based decisions need to be explainable in order to be trusted.

There is a clear need, therefore, for those in the C-suite to review the AI practices within their companies, ask a series of key questions, and—where necessary—take steps to tackle a variety of potential risks from AI, by addressing any areas where controls or processes are found to be lacking or inadequate. The risks include those related to biased decision-making, the interpretability of AI decisions, a lack of explainability and the possibility that AI-powered systems could displace human workers. Other risks include higher-level societal concerns about how AI could amplify inequality between the haves and have-nots—or even pose physical threats to humans, ranging from individual injury to mass annihilation by autonomous weapons.








# Responding to AI challenges across 5 key dimensions



Alongside these risks, the rise of AI also brings inherent challenges around trust and accountability. To tackle these effectively, organisations should both understand the challenges and risks around AI and take these fully into account in its design and deployment.

PwC has developed a comprehensive Responsible AI Framework and Toolkit to help companies focus on and address five key dimensions when designing and deploying responsible AI applications:

- 1**  **Governance**  
Governance serves as an end-to-end foundation for all the other dimensions.
- 2**  **Ethics and regulation**  
The core goal is to help organisations develop AI that is not only compliant with applicable regulations, but is also ethical.
- 3**  **Interpretability and explainability**  
Provides an approach and utilities for AI-driven decisions to be interpretable and easily explainable by those who operate them and those who are affected by them.
- 4**  **Robustness and security**  
Helps organisations develop AI systems that provide robust performance and are safe to use by minimising the negative impact.
- 5**  **Bias and fairness**  
Addresses the issues of bias and fairness—recognising that while there is no such thing as a decision that is fair to all parties, it is possible for organisations to design AI systems to mitigate unwanted bias and achieve decisions that are fair under a specific and clearly-communicated definition.



The foundation for Responsible AI is end-to-end enterprise governance. At its highest level, AI governance should enable an organisation to answer critical questions about results and decision-making of AI applications, including:

- Who is accountable?
- How does AI align with the business strategy?
- What processes could be modified to improve the outputs?
- What controls need to be in place to track performance and pinpoint problems?
- Are the results consistent and reproducible?

**Historically, governance functions have only had to deal with static processes. But AI processes may be dynamic and adaptive —and AI governance must be as well.**

The ability to answer such questions and respond to the outcomes of an AI system requires a more flexible and adaptable form of governance than many organisations may be accustomed to.

Historically, governance functions have only had to deal with static processes. But AI processes are iterative —and AI governance must be as well. A proper AI governance foundation will start with strategy and planning across the organisation, but will also take into account existing capabilities and the vendor ecosystem, as well as the unique model development process and model monitoring and compliance.



# PwC's enterprise governance framework covering the AI lifecycle





Organisations should strive to develop, implement, and use AI solutions that are both morally responsible and also legal and ethically defensible. More than 70 documents have been published in recent years to describe relevant ethical principles for AI.

While it is hard to dispute the ethical principles, businesses find it challenging to translate them into concrete actions that impact day-to-day decisions. For principles to become actionable, they must be contextualised into specific guidelines for front-line staff. By contextualising the ethical considerations, organisations may be able to identify the ethical implications of their AI solutions, and the relevant principles that should be taken into account when designing and operationalising AI models, allowing for robust mitigation of ethical risks.

Organisations also must monitor the regulatory environment in which they operate and understand how emerging regulations will shape future business practices.



At some point, any business using AI will need to explain to various stakeholders why a particular AI model reached a particular decision. These explanations should be tailored to the different stakeholders, including regulators, data scientists, business sponsors, and end consumers.

A lack of interpretability in AI decisions is not only frustrating for end-users or customers, but can also expose an organisation to operational, reputational, and financial risks.

A lack of interpretability in AI decisions is not only frustrating for end-users or customers, but can also expose an organisation to operational, reputational, and financial risks. To instill trust in AI systems, people must be enabled to look “under the hood” at their underlying models, explore the data used to train them, expose the reasoning behind each decision, and provide coherent explanations to all stakeholders in a timely manner.





To be effective and reliable, AI systems need to be resilient, secure, and safe. In terms of resilience, next-generation AI systems are likely to be increasingly “self-aware,” with a built-in ability to detect and correct faults and inaccurate or unethical decisions. In terms of security, the potentially catastrophic outcomes of AI data or systems being compromised or “hijacked” make it imperative to build security into the AI development process from the start, being sure to cover all AI systems, data, and communications. For example, if the image of a ‘STOP’ sign is manipulated to be misinterpreted as a ‘30 MPH’ speed limit sign, this could potentially result in a disastrous situation for the passengers (or pedestrians) of an autonomous vehicle.

Above all, though, AI systems must be safe for the people whose lives they affect, whether they are users of AI or the subjects of AI-enabled decisions. This is clearly crucial in areas such as healthcare, autonomous vehicles, and connected worker or manufacturing applications.





Bias is often identified as one of the biggest risks associated with AI, as underlined by recently reported cases of known bias in AI, such as racial bias in criminal justice systems<sup>2</sup> and gender discrimination in hiring<sup>3</sup>. Eliminating bias is a more complex task than it may appear, however.

The public discussion about bias in AI often assigns blame to the algorithm itself, without taking the human component into account. And people perceive bias through the subjective lens of fairness—a social construct with strong local nuances and many different and even conflicting definitions. In fact, it's impossible for every decision to be fair to all parties, whether AI is involved or not. But it is possible to tune AI systems to mitigate bias and enable decisions that are as fair as possible and adhere to an organisation's corporate code of ethics, as well as following anti-discrimination regulations.

That said, even having clear criteria doesn't remove the need to make trade-offs between competing—and sometimes conflicting—priorities. In this sense, AI decisions are similar to those made by humans. In each case, establishing fairness requires businesses to choose their level of comfort in the choices they make, and balance these against the associated costs and wider impacts, which might be negative for some stakeholders.

AI decisions are similar to those made by humans. In each case, establishing fairness requires businesses to choose their level of comfort in the choices they make, and balance these against the associated costs and wider impacts, which might be negative for some stakeholders.

<sup>2</sup> [Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks. ProPublica, 23 May, 2016](#)

<sup>3</sup> [Amazon scraps secret AI recruiting tool that showed bias against women. Reuters, 10 October 2018](#)



# PwC's Responsible AI Toolkit





PwC's Responsible AI Toolkit consists of a flexible and scalable suite of capabilities, covering frameworks and leading practices, assessments, technology, and people. The Responsible AI Toolkit is designed to enable and support the assessment and development of AI across an organisation, tailored to its unique business requirements and level of AI maturity.

The Toolkit enables organisations to build high-quality, transparent, explainable and ethical AI applications that generate trust and inspire confidence among employees, customers, business partners, and other stakeholders.

PwC's Responsible AI Toolkit enables organisations to build high-quality, transparent, explainable and ethical AI applications that generate trust and inspire confidence

### **Frameworks and leading practices**

The frameworks, leading practices, and practice aids in the Toolkit help organisations to define the requirements for Responsible AI governance.

For example, the Toolkit includes an Ethical AI Framework that helps organisations contextualise and apply ethical principles in practice. The framework differs from other comparable approaches because it provides customised, robust methodology for making technical decisions, identifying, contextualising, and mitigating ethical risks.

## **Assessments**

There is often a disconnect between business expectations and current-state performance. The Toolkit's diagnostics are designed to help different stakeholders within the business make comparisons, identify gaps, and strengthen overall processes.

## **Technology**

It's important to recognise how heavily machine learning models rely on large quantities of data. In most cases, the data needs to be labeled, which is a significant and time-consuming task. If the data is biased, the model will be biased. The data should be used in the appropriate context, and the applicable techniques should be used to enable the trade-off between accuracy, explainability, fairness, and security.

The Toolkit is enabled by technology, providing a set of assets that are curated to accelerate the evaluation of data, models and their trade-offs, taking into account the relevance and associated risks, and the AI maturity of an organisation.

Toolkit accelerators include the ability to rapidly assess whether the outputs from a particular model fall under different definitions of bias. Users are then able to evaluate the trade-offs around performance and cost from making incremental improvements to the system.

## **People**

Within any organisation, there are different stakeholders involved in the use of AI, such as business sponsors, process owners, data scientists, and end consumers. The Toolkit provides the recommended level of interaction with each of these stakeholders to help create trust.







# The business imperative of Responsible AI

For any organisation to realise the full promise of AI, it must address the dimensions described above. In today's increasingly transparent, fast-moving and competitive marketplaces, implementing Responsible AI is not merely a nice-to-have, but a prerequisite for success.

Put simply, if AI isn't responsible, it isn't truly intelligent. Companies must bear this in mind as they plan and build their AI-enabled future. PwC can help.

```
elif _operation == "MIRROR_Y":  
    mirror_mod.use_x = False  
    mirror_mod.use_y = True  
    mirror_mod.use_z = False  
elif _operation == "MIRROR_Z":  
    mirror_mod.use_x = False  
    mirror_mod.use_y = False  
    mirror_mod.use_z = True
```

```
#selection at the end -add back the deselected mirror modifier object  
mirror_ob.select= 1  
modifier_ob.select=1  
bpy.context.scene.objects.active = modifier_ob  
print("Selected" + str(modifier_ob)) # modifier ob is the active ob  
#mirror_ob.select = 0  
#name = bpy.context.selected_objects[0]  
#bpy.data.objects[name].select = 1
```

# Contacts

## **Anand Rao,**

Global Artificial Intelligence Leader, PwC US  
anand.s.rao@pwc.com

## **Flavio Palaci,**

Global Data & Analytics Leader, PwC Australia  
flavio.j.palaci@pwc.com

## **Wilson Chow,**

Global and Mainland China/Hong Kong Technology, Media and  
Telecommunications Industry Leader, PwC China  
wilson.wy.chow@cn.pwc.com

To learn more, take PwC's free Responsible AI Diagnostic.



Or check our Responsible AI website.

[www.pwc.com/rai](http://www.pwc.com/rai)



© 2019 PwC. All rights reserved. PwC refers to the PwC network and/or one or more of its member firms, each of which is a separate legal entity. Please see [www.pwc.com/structure](http://www.pwc.com/structure) for further details. This content is for general information purposes only, and should not be used as a substitute for consultation with professional advisors.