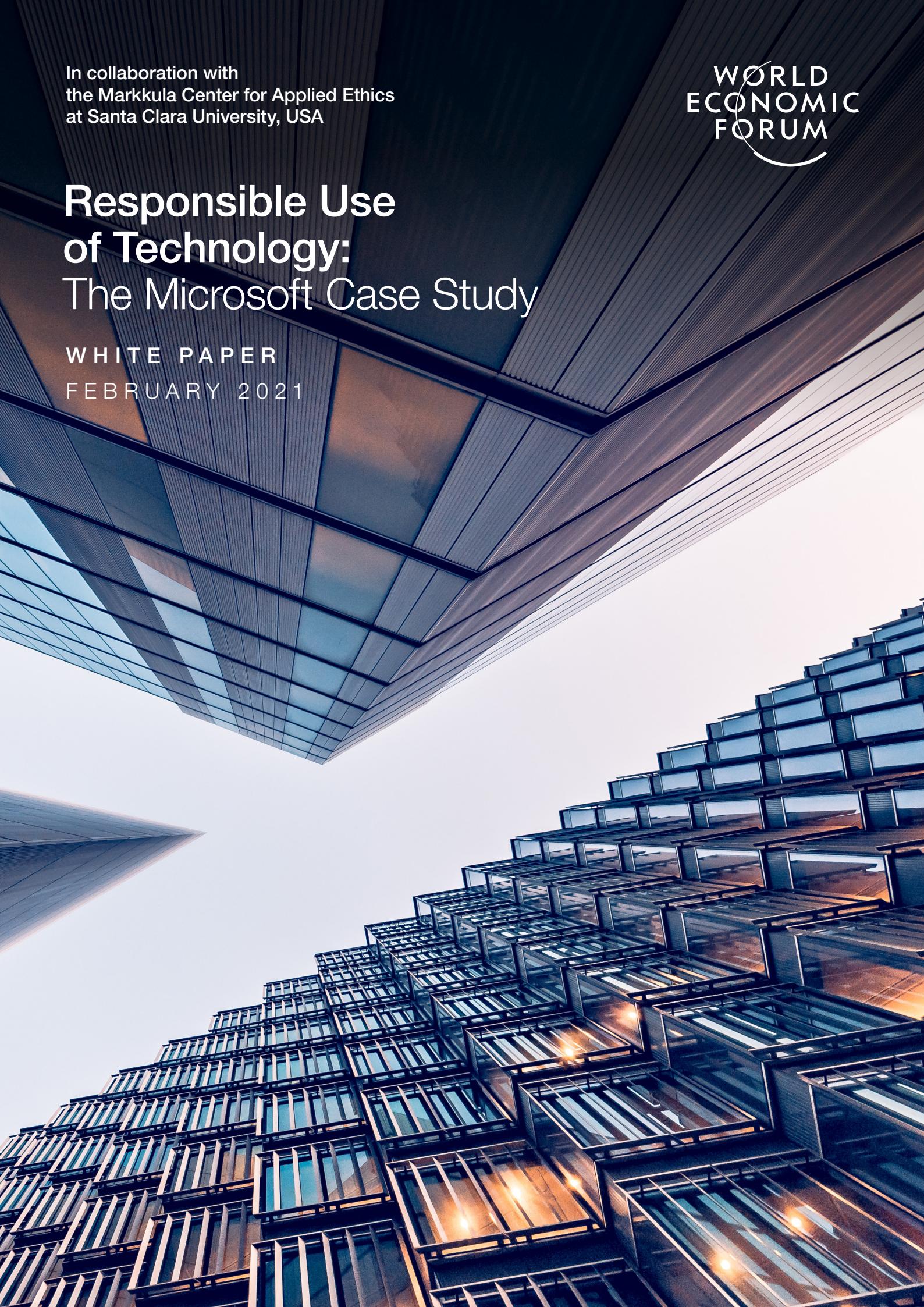


In collaboration with
the Markkula Center for Applied Ethics
at Santa Clara University, USA



Responsible Use of Technology: The Microsoft Case Study

WHITE PAPER
FEBRUARY 2021



Contents

3	Foreword
4	1 Introduction and challenge
5	2 The Microsoft experience
6	3 How to change a culture
6	3.1 Governance
7	4 Microsoft AI principles
8	5 Responsible AI Standard and processes
9	6 Tools for responsible innovation
9	6.1 Judgment call
10	6.2 Envision AI workshop
10	6.3 Impact assessment
11	6.4 Community Jury
11	6.5 Machine learning tools with ethical impact
15	7 Measuring culture change
16	8 Analysis of best practices
17	9 What remains to be done
18	Conclusion
18	Box 1 Project Tokyo: Applying responsible innovation practices
20	Contributors
21	Endnotes

© 2021 World Economic Forum. All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, including photocopying and recording, or by any information storage and retrieval system.

Foreword



Brian Green,
Director, Technology
Ethics, Markkula Center for
Applied Ethics, Santa Clara
University, USA



Don Heider,
Executive Director, Markkula
Center for Applied Ethics,
Santa Clara University, USA



Kay Firth-Butterfield,
Head, Artificial Intelligence
and Machine Learning;
Member of the Executive
Committee, World Economic
Forum LLC



Daniel Lim,
Project Fellow, Artificial
Intelligence and Machine
Learning, World Economic
Forum LLC; Seconded
from Salesforce

The World Economic Forum Centre for the Fourth Industrial Revolution was launched in 2017 with the mandate to co-create policy and governance frameworks through a multistakeholder approach to accelerate the adoption of emerging technologies. The Centre's platforms include areas such as artificial intelligence and machine learning, blockchain, data policy and internet of things. At the heart of this work is the drive towards action, transparency, ethics and global public good.

Today, the global Coronavirus pandemic continues to cast a dark shadow over all facets of society. As social distancing measures became a necessity to preserve public health, digital transformation became a requirement for most businesses to simply survive. The urgency to maximize the benefits of technology, while mitigating the risks and harms, has never been greater. It is incumbent on all organizations that design, develop, procure, deploy and use technology to do so in a responsible manner.

In our numerous conversations with leaders across the various sectors, we've learned that a gap exists between organizations' desire to act ethically and their understanding of how to follow through on their good intentions. We refer to this as an intention-action gap. To this end, the Centre is focused on providing practical resources for organizations to operationalize ethics in technology. This initiative, which began in 2019, with active participation from civil society, governments and companies, made the case for both human-rights-based and ethics-based approaches to the [responsible use of technology](#). To help bridge this

intention-action gap, we aim to provide leaders with practical tools for how they might: 1) educate and train their employees to think more about responsible technology; 2) design their organization to promote more ethical behaviour and outcomes; and 3) design and develop more responsible technology products.

It is with this last goal in mind that the World Economic Forum and the Markkula Center for Applied Ethics at Santa Clara University publish this White Paper, the first in a series of case studies highlighting tools and processes that facilitate responsible technology product design and development. This initial document on the "Responsible Use of Technology: The Microsoft Case Study" will be followed by other companies' examples of ethical practices and tools in future papers. We thank Microsoft for having shared their responsible innovation tools, practices and expertise for this effort. It is our hope that this document will inspire others to contribute to the Forum's [Responsible Use of Technology project](#) by sharing tools and methods that businesses have created for the same purpose.

To achieve these ambitious goals requires the collaboration of all global stakeholders. The World Economic Forum, committed to improving the state of the world, is the International Organization for Public-Private Cooperation. The Markkula Center for Applied Ethics at Santa Clara University in California, a key partner in this project, has over 30 years' history and experience in promoting ethical deliberations. Together, we are pleased to collaborate towards this ambitious vision.

Introduction and challenge

Our society is undergoing a Fourth Industrial Revolution,¹ in which powerful technologies such as artificial intelligence (AI), the internet of things (IoT) and augmented reality have the potential to create lasting societal benefits. However, without the right guard rails, these technologies can also cause immense harm. Organizations have a responsibility to design, develop, procure, deploy and use technologies in a responsible manner. As an initial step towards more responsible innovation, dozens of groups from industry, civil society and government have published ethical technology principles, particularly on AI.² These values and frameworks offer a foundation for what responsible technology innovation outcomes might look like. Yet the need to define *how* technology products can be ethically made remains.³

This paper is the first in a series of case studies that investigate how companies have begun to incorporate ethical thinking into the development of technology. It focuses on Microsoft Corporation and will be followed by papers describing efforts in other companies. Through a series of interviews with Microsoft executives and employees, combined with secondary research, this paper presents an outside perspective of Microsoft's current and evolving efforts to build on the ethical values and culture of the company with tools and practices in its product engineering organization.

The effort to operationalize ethics in technology for any organization should be a continuous journey with companies always looking for ways to improve. Microsoft's experience is no exception. The purpose of this paper and this series, in the context of the World Economic Forum's Responsible Use of Technology project, is to surface lessons that can help organizations advance their own responsible innovation practices. The project aims to identify areas for improvement in tools and processes that can help drive more ethical considerations in technology product development. It may even inspire others who have created new methods for this cause to share their work, either in this series or elsewhere.

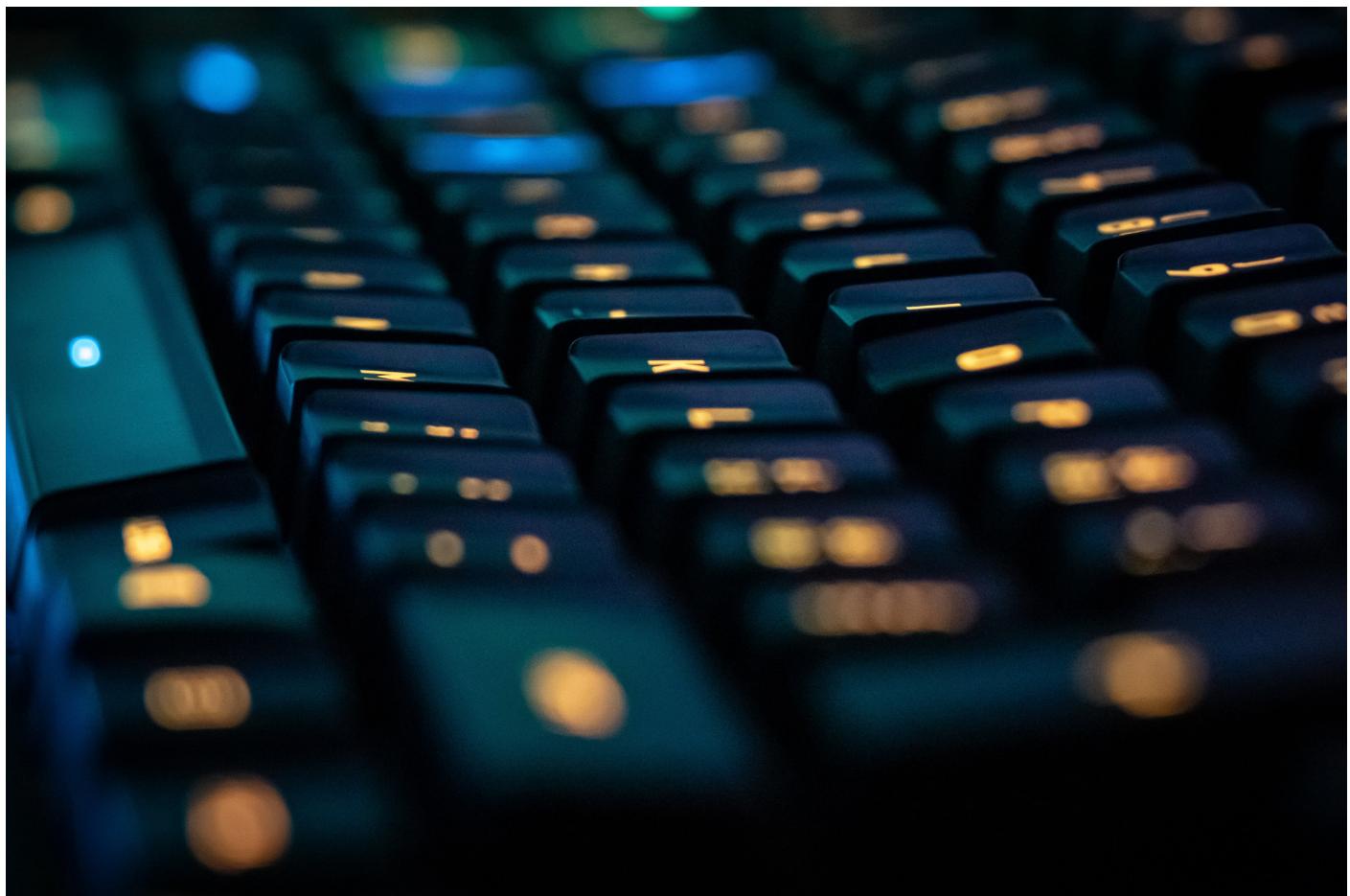
The Microsoft experience

When Satya Nadella became Chief Executive Officer of Microsoft in 2014, he brought with him the concept of the “growth mindset”. This concept originated from the work of Stanford University professor Carol Dweck, whose research led her to discover the powerful idea of *mindsets*. Dweck’s work demonstrates that success in human endeavours can be highly influenced by how one thinks. Someone who has a fixed mindset is less likely to succeed than someone with a growth mindset.⁴ From the beginning of his tenure as CEO, Nadella instilled the growth mindset into Microsoft’s culture. This mindset encourages curiosity, experimentation, hard work and learning, and repeating the process. This evolution in Microsoft’s culture created a corporate environment that promoted innovation and introspection on the impact of technology on society.

At every level, people at Microsoft already encounter and respond to ethical questions in their work. One notable example is the release of a chatbot named Tay in 2016, an experiment on conversational

understanding. The hope was that the more Tay interacted with humans in conversations on Twitter, the “smarter” the AI-powered bot would get. However, soon after its debut, a group of users maliciously targeted Tay, causing Tay to respond with inappropriate and denigrating responses. Microsoft quickly withdrew Tay from the public and issued an apology.⁵ This episode, as well as others, was pivotal in motivating the organization to incorporate ethical considerations into its product innovation process.

Technologies are powerful, and that power can be used for good or for ill. The designers of new technologies must therefore think carefully about how technologies are built and help to increase the likelihood that they will be used for the benefit of society. Microsoft is one company that is taking steps to develop quality products that empower customers to achieve more in a way that also allows them to be used responsibly.



How to change a culture

Driven by a commitment to strengthen trust with customers and society more broadly, the leadership and employees of Microsoft have taken a values-based approach to building and deploying technology responsibly. While this approach is certainly more demanding in terms of time, money and effort, the company believes it has the best return on investment over the long term.

What motivated Microsoft to take this approach to solving this challenge? Satya Nadella's 2016 "Partnership of the Future"⁶ article in *Slate*'s "Future

Tense" describes it well: this decision to make socially aware and ethical technology is not merely about profit but about the world created for humans to live in. Technology developers are not only technologists. They are human beings who must live in the world that technology creates. Keeping this fact in the foreground ensures that the bigger issues of the social and environmental impact of technology remain close at hand, and directs technologists towards developing more humane technologies that can truly benefit the common good of all humankind and the planet.



Building trust in technology is crucial ... it starts with us taking accountability ... for the algorithms we create, the experiences we create, and ensuring there is more trust in technology with each day.

Satya Nadella, Chief Executive Officer, Microsoft⁷

Importantly, Microsoft's choice to emphasize ethics was not only a top-down decision but one that emerged within the company at many levels. The company's approach, which is based on its AI principles (see the next section), focuses on proactively establishing guard rails for AI systems that will ensure that potential risks are anticipated

and mitigated, while maximizing benefits to society. The building blocks for this responsible AI programme include a governance structure to ensure accountability and enable progress; a set of rules defining its Responsible AI Standard; and training, tools and practices that allow employees to operationalize the principles and rules.

3.1 Governance

Microsoft's AI governance approach follows a "hub-and-spoke" model that helps the company integrate privacy, security and accessibility into its products and services. Three teams play critical roles in this governance approach. First, the Aether Committee comprises working groups of scientific and engineering experts to advise on responsible AI issues and the enactment of the company's AI principles. Second, the Office of Responsible AI manages the policy, governance, enablement and sensitive uses functions. Third, the Responsible AI Strategy in Engineering (RAISE) team enables Microsoft's engineer groups to implement responsible AI processes through the adoption of systems and tools.

As for the "spokes" in this governance approach, Microsoft has found consistent success in deploying a "Champs" model, in which respected domain experts across teams and regions are appointed by leadership

to promote awareness and education on a given subject. The Responsible AI Champs programme was implemented in 2020 and includes experts in the areas of security and open source, among others. Champs bring attention to the available responsible AI tools and processes. They also help teams identify and consider ethical and societal issues in their work. They have been crucial for leading a culture shift towards thinking more about ethics.⁸

Another important spoke in the governance approach is Ethics & Society, a team within RAISE tasked with taking a design thinking approach to ethics. It has played a key role in the creation of some of the responsible innovation tools described in this paper.⁹

Shifting the culture of a corporation is a monumental feat, but it is certainly not impossible, and the resources in industry, civil society and academia for thinking about these efforts are excellent.¹⁰



The most critical next step in our pursuit of AI is to agree on an ethical and empathic framework for its design.

Satya Nadella, Chief Executive Officer, Microsoft¹¹

Microsoft AI principles

To promote a culture shift and infuse its AI work with ethical awareness, Microsoft developed six overarching ethical principles for AI: fairness, reliability and safety, privacy and security, inclusiveness, transparency and accountability.¹²

Fairness	Reliability and safety	Privacy and security
AI systems should treat people fairly	AI systems should perform reliably and safely	AI systems should be secure and respect privacy
Inclusiveness	Transparency	Accountability
AI systems should empower and engage people	AI systems should be understandable	People should be accountable for AI systems

These six principles act as a mental tool or framework in which to organize thinking about ethics at Microsoft. While specifically phrased as responsible AI principles, they have relevance for much of the work in the technology industry. A brief explanation of each principle follows:

Fairness

Focuses on developing systems that treat everyone in a fair and balanced way. This principle acknowledges that defining and mitigating fairness issues for a system depend on understanding the system's purpose and context of use, and that a system's fairness reflects decision-making during both development and deployment.

Reliability and safety

Means developing systems that are robust and capable of maintaining safe operations even in worst-case scenarios. This principle encompasses consideration of the harms that might come from a technology, and ways employees can strive to minimize those risks, so technologies can give the greatest benefits to their users.

Privacy and security

Seeks to protect data and use data in a way that is secure for all stakeholders. Privacy is a basic right and protecting it is crucial for ensuring that stakeholders can trust companies with their data. Data must be secure at all stages and, to further this end, actions must be taken to institutionalize privacy and security for the data companies are responsible for.

Inclusiveness

Makes sure that no one is left out of the design, development, deployment and use of technology. Communities across the full spectrum of humanity should be meaningfully engaged and empowered by technology, and technology should not be limited to only a few privileged communities. This inclusion should not only involve building for, but building with, the diverse stakeholders.

Transparency

Seeks to create technology that is intelligible and explainable, not only to those who are developing the technology but also to those who will be using it, or will be affected by it. Stakeholders should be able to interpret and understand what a technology is doing and why it is acting that way. This allows product teams to contextualize and improve results.

Accountability

Means that people take responsibility for the way technology operates and for the impact of that operation on society. This includes considering the structures that can be implemented to ensure accountability at multiple levels, including design, development, sales, marketing and use, as well as advocacy for the regulation of technologies when warranted.

Simply having principles does not change a culture unless those principles are made concrete through tools and practices that help employees work through how to think ethically. To advance these principles and make sure they are implemented into the company's workflows, Microsoft developed several tools for incorporating applied ethics in technology. All of these tools serve an ethical end; some are more procedural, while others are more technical in nature.

Responsible AI Standard and processes

With the goal of turning principles into practice, Microsoft created the Responsible AI Standard in 2019 as part of a process of learning how best to enact Microsoft's AI principles across the company. The standard outlines a set of steps that teams must follow to support the design and development of responsible AI systems. A key part of the standard is a set of responsible AI considerations with examples that guide teams through the AI system development life cycle. The standard has been piloted with 10 internal engineering groups and two customer-facing groups. Although teams found the responsible AI considerations and examples helpful, they requested more specificity on the requirements and criteria they could apply to their situations. Based on this feedback, version 2.0 of the standard was developed and is now being previewed with employees ahead of its roll-out across the company more broadly. This version

will reinforce a human-centred approach, as well as strong engineering and research foundations. For each requirement in the 2.0 Standard, Microsoft is building implementation methods that teams can follow to support the creation of responsible AI systems.

One of the Responsible AI Standard's requirements is for sensitive use cases that meet predefined review criteria to be reported and escalated to the Office of Responsible AI. If there is no previous precedent to draw upon, members of that office participate in a process of deliberation.

In 2019, Microsoft also created the Introduction to Responsible AI training course for staff, which covers the sensitive use process, the Responsible AI Standard and the foundations of its AI principles. The training is now mandatory for all employees.



Tools for responsible innovation

Technology ethics is fundamentally an applied, practical discipline. It is not about theory, although theory informs practice. Technology ethics, as a practical pursuit, benefits from having very specific and applicable tools to aid thinking, analysis, stakeholder engagement and decision-making. The tools presented here – Judgment Call, Envision AI workshops, impact assessments, Community Jury, Fairlearn, InterpretML and the error terrain analysis – each address a specific aspect of the ethics of technology development. Most of these tools, guidelines and resources are available to the public

at [Microsoft's Responsible AI Resource Center](#). Together, institutionalized into a structure and applied systematically, they can help to improve products and prevent the potentially damaging effects of technologies. Applying these tools, regularly and routinely, can help to shift a culture, especially in conjunction with techniques such as securing leadership buy-in, establishing prosocial norms and implementing ethical reminders.¹³ These responsible innovation tools have been applied to improve Microsoft technologies, such as spatial analysis, speech consent and Custom Neural Voice.¹⁴

6.1 Judgment call



Source: Microsoft

To help cultivate empathy during its product creation process, Microsoft's Ethics & Society team created the Judgment Call game.¹⁵ The game is an interactive team-based activity that puts Microsoft's AI principles of fairness, privacy and security, reliability and safety, transparency, inclusiveness and accountability into action. During the game, each participant is given a card that assigns them a role as an impacted stakeholder of a digital product (e.g. product manager, engineer, consumer). Each is also given a card that represents one of Microsoft's AI principles, and a card with a number from 1 to 5, representing the stars in a ratings review. Participants are asked to write a review of the digital product from the perspective of their assigned role, principle and rating number. Each player is asked to share and discuss their review.

The game has a number of benefits:

- Engineers, product managers, designers and technology executives consider the perspectives of the impacted stakeholders and imagine the potential outcomes of their product on these stakeholders.
- Although the game does not replace the valuable benefits of interacting directly with stakeholders, it builds empathy, especially early in the product design process.
- Roles are arbitrarily assigned to participants due to the random distribution of the cards. The game's dynamics create a safe environment for product team members to discuss potentially sensitive ethical topics.

6.2 Envision AI workshop



Source: Microsoft

Developed by the Project Tokyo team (see the Project Tokyo text box below) from Microsoft Research, Engineering Learning & Insights and the Office of Responsible AI, the Envision AI workshop is an exercise that educates Microsoft teams on how to conduct an impact assessment, a process required in the Responsible AI Standard. In an interactive and engaging setting, Envision AI participants examine real scenarios that occurred while developing the assistive AI system in Project

Tokyo. They learn the human-centric design approach to AI and the resources available to them to identify the potential effects of the technology on the stakeholders. Participants apply these lessons to completing an impact assessment. The Envision AI workshop helps Microsoft empower teams to conduct ethical deliberations on their own and take responsibility for the implications of the products they create.

6.3 Impact assessment

Conducting an impact assessment is a required step in the product development process of all AI projects at Microsoft. Teams complete an extensive questionnaire, which takes into account the intended use cases of a product and its potential impacts on stakeholders, and a self-assessment of the potential risks. The completed impact

assessments are reviewed by peers and executives at the company. This process is facilitated and required by the company's Office of Responsible AI. It serves as an important tool to help ensure the responsible development and deployment of AI across Microsoft.

6.4 | Community Jury



Source: Center for New Democratic Processes, USA

Community Jury is a technique that allows project teams to directly interact with impacted stakeholders.¹⁶ A group of representative stakeholders from diverse backgrounds are recruited and selected to be jury members. During a Community Jury session, project teams provide the jury members with an overview of the product's purpose and its potential use cases, benefits and harms. Participants share information and discuss their perspectives of the product's impacts with the facilitation of a neutral moderator. As key themes emerge from the discussion, the participants jointly define the opportunities and challenges presented by the technology. This process can also lead to co-created solutions.

The planning process starts by aligning goals and outcomes with the project teams. It is important that product teams allocate time in the product development process to conduct Community Jury sessions. Based on the project objectives, jury recruitment and selection should be diverse and inclusive. Strong session facilitation by a neutral

moderator is important to allow every voice to be heard. Moderators need to provide ample time for knowledge sharing, deliberation and co-creation. Finally, it is important to disseminate a report on the Community Jury outcome that summarizes the key insights for transparency.

The benefits of the Community Jury technique from an ethical perspective are multifaceted. Product teams and impacted stakeholders are brought together to learn from each other's perspectives. The proximity and connection helps build community and empathy. Especially for product teams, stakeholder engagements raise their awareness of issues otherwise not readily apparent when the new technologies were conceived. This process can build consensus among teams and their communities on the challenges and opportunities that technological innovations may pose. It also presents a vital opportunity for teams to improve their products and solutions to benefit a larger group of stakeholders.

6.5 | Machine learning tools with ethical impact

Some of Microsoft's tools for considering ethics are technical devices to understand, assess and mitigate the ethical risks of machine learning models. They serve multiple ethical AI principles – namely, that AI must be fair, reliable, inclusive, transparent and accountable. These software tools are constantly being developed, refined and changed.

Fairlearn

Fairlearn is an open-source toolkit designed for data scientists, developers, business stakeholders

and researchers to help them assess and improve fairness in machine learning.¹⁷ Fairlearn has two main components: 1) a set of fairness assessment metrics and an interactive data visualization dashboard, which provide an understanding of how particular groups may be adversely affected by models (this allows a comparison of fairness and performance metrics between models); and 2) unfairness mitigation algorithms for a variety of AI tasks, as well as definitions of fairness to allow deeper thought in this context.¹⁸

FIGURE 1

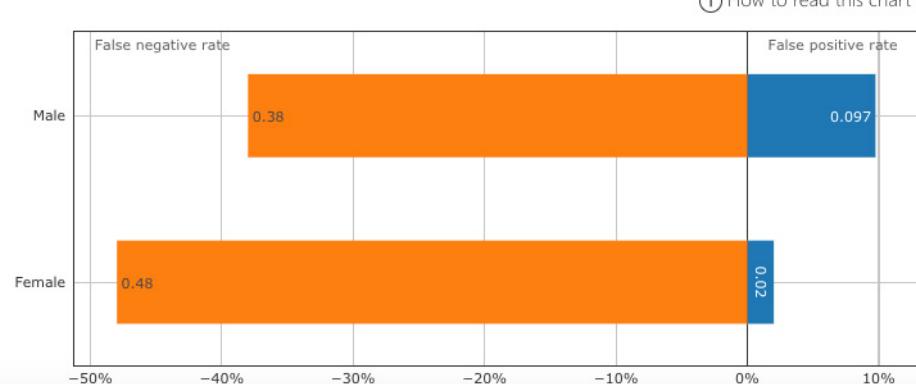
Fairlearn tool dashboard

Assessment results for Model 0

Sensitive feature	Performance metric	Fairness metric
sex	Accuracy	Demographic parity difference
	Accuracy	Selection rate
Overall	85.4%	19.6%
Male	81.6%	25.7%
Female	93.1%	7.36%

Charts

False positive and false negative rates



Source: Microsoft

Fairlearn's creators note that the reasons models can behave unfairly are many, including social, technical and combined sources of unfairness.¹⁹ Fairlearn focuses on the negative impacts of models on groups of people, such as those defined in terms of race, gender, age or disability status. It contains understanding and mitigation tools to assess quality of service, such as, for example, whether a voice recognition system works as well for one group as for another. The tools can also assess allocation, for instance whether a system that identifies job applicants unfairly recommends applicants of one ethnicity over another.

Since fairness is a deeply sociotechnical concept, using Fairlearn does not guarantee that a model can be made perfectly fair. It only promises to help identify and mitigate fairness issues involving harms of allocation and harms in quality of service. The authors of the tool rightly note that perfect fairness is not possible, not only because of the nature of data from the real world but also because of the theoretical incompatibilities of some definitions of fairness. Importantly, Fairlearn is specifically aimed at mitigating fairness-related harms towards protected groups (e.g. different ethnicities).²⁰ In one case, Fairlearn significantly improved fairness for loan decisions.²¹

InterpretML

Microsoft developed the InterpretML open-source toolkit to make machine learning models more transparent, intelligible and interpretable.²² It can help to provide both "global" explanations about overall model behaviour and "local" explanations for individual model predictions. InterpretML contains "glass box" (or inherently explainable) models, including explainable boosting machines and decision trees, as well as a number of tools that help to explain "black box" models. The toolkit also supports "what-if" explanations for model outputs; most recently, it also added "diverse counterfactual explanations", which compute the most similar data instances that have received different prediction outcomes. A counterfactual analysis generates explanations for individual outputs or predictions by identifying the smallest change to the input features that would cause the model or system to produce a desired output or prediction.

Transparent, intelligible and interpretable models are desirable because they make models much easier to understand and debug. They also provide advantages that include making the model easier to explain to others, such as end users, helping to discover sources of fairness issues and clarifying options for informing unfairness mitigation techniques, and making models clear for compliance with relevant regulatory obligations.

These technical improvements should be recognized for what they are: ethical improvements. For example, buggy software can be potentially dangerous and harmful in numerous ways. These harms are ethically significant, and being able to remove them ultimately contributes to a more ethical product. The opportunity to discover and mitigate the fairness issues of a model is also a significant ethical benefit, both for recognizing where past wrongs may have been systemic (and, for these, a historical wrongs recompense may be in order) and for mitigating these wrongs. And prioritizing the clear and timely fulfilment of regulatory obligations can greatly simplify processes in the long term, saving labour and money that could be better directed elsewhere.

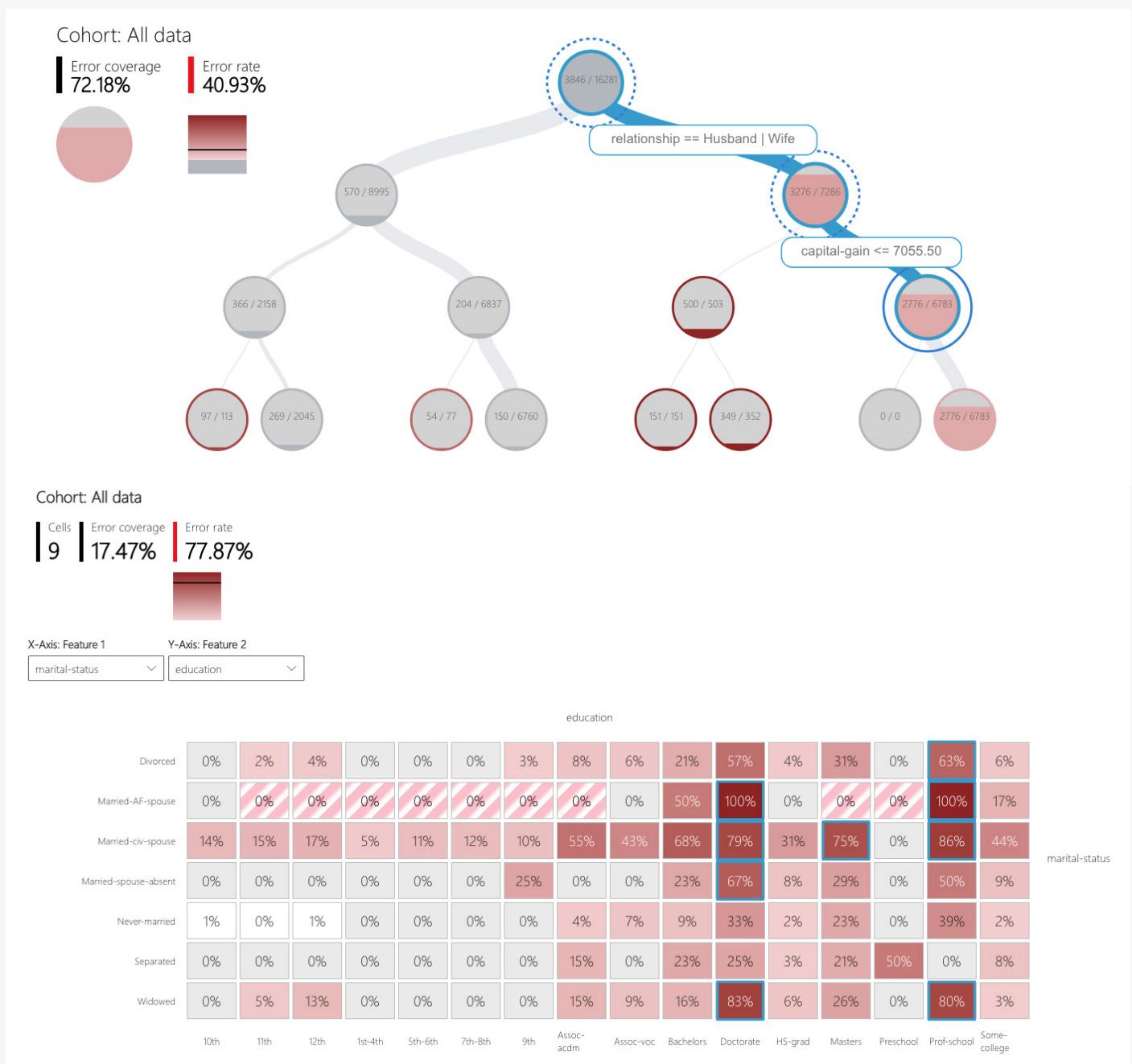
Overall, making machine learning more transparent helps to make it more trustworthy, and builds trust and reputability.

Error terrain analysis for machine learning

The error terrain analysis for machine learning tool (often called simply the Error Analysis tool) is similar to InterpretML in that it works to debug machine learning models. However, its specific role is to help debug exactly those classifications within a model that might be subject to errors.²³ The tool enables data scientists to identify cohorts of data with high error rates versus the benchmark rate and to visualize how the error rate distributes. Via integration with interpretability techniques and visualizations, users can further diagnose the root causes of the errors by gaining deep insights into the data or model.

FIGURE 2

Error Analysis tool data displays



Source: Tool images provided by Microsoft

The Error Analysis tool solves a core technical challenge in machine learning, which is to determine exactly where a model misclassification occurs. For example, in a dataset of faces for a facial recognition model, the tool might discover that skin tone correlates with misclassification of sex, or that beards or hair colour are a cause of various other misclassifications. By pinpointing exactly where these misclassifications occur, developers can determine how to fix the model to improve its accuracy.

The Error Analysis tool enables machine learning practitioners to efficiently analyse and debug machine learning models' errors to accelerate improvements in model accuracy and reliability. They can discover new insights about relationships between data features and model performance, dive deeper beyond overall performance metrics by performing a disaggregated evaluation, and discover how errors are distributed in various cohorts. They can further leverage this knowledge to deploy mitigation techniques and improve the model or the data.

As with InterpretML, the Error Analysis tool could be construed as simply a technical fix in order to gain accuracy in machine learning development. But in certain cases, such as facial recognition, failures in classification are ethically significant. For example, inaccuracies with respect to race and sex can both reflect deep fairness issues in society and perpetuate injustice at a systemic level, not to mention their being offensive. And other sorts of inaccuracies in processing visual data – for example with autonomous vehicles – could cause safety concerns.

Once again, technical sophistication and ethics go hand in hand. Ethical mistakes caused by technical errors can be prevented by eliminating these errors from products. As one Microsoft interviewee stated, "We don't use our customers as testers!" Instead, the company strives to test and adjust a product thoroughly before it ever gets to the customer (though, of course, it does receive customer feedback on improvements that remain to be made, in the form of comments, press, error reports, etc.).

Measuring culture change

Periodic goal setting and performance evaluations are essential exercises to build alignment and measure progress at any organization. Microsoft has extended these organizational techniques to include considerations of ethics and responsible behaviours.

In 2018, a product team under the leadership of Alex Kipman (inventor of Kinect and the HoloLens) was working on reducing error rates in facial recognition technology at Microsoft. Inspired by this work, Kipman's colleagues in the Ethics & Society team suggested to management that the facial recognition project team add a "core priority" related to reducing bias to each team member's annual performance goal. They also suggested that the team commit to external audits. This biannual goal setting and individual performance review exercise at Microsoft is called a "Connect". Kipman championed this idea, expanding the practice beyond his facial recognition team to his entire organization, which included all the AI cognitive services and mixed reality teams at Microsoft. He supported an 18-month culture change process to incorporate this new responsible AI and ethics core priority into each employee's performance review. Every employee was evaluated twice a year, in part on their responsible AI commitments and how they implemented them. Because this was a significant new individual requirement and organizational commitment, a virtual team was formed with representation from across Kipman's organization (e.g. Human Resources, Engineering, Communications) to work on a culture change plan and roll-out strategy. The Ethics & Society team, which initiated this programme, launched an educational initiative consisting of workshops, informational sessions, presentations and pamphlets with the goal of educating Kipman's 1,400-person team on how

to write coherent, individualized core priority statements and implement the company's responsible AI work successfully.

Microsoft's Ethics & Society team learned some important lessons from this effort. First, for over three months, every person on the team was tapped to participate in the delivery of weekly workshops and informational sessions. Their purpose was to help each individual employee understand why the new responsible AI commitment existed and how it was related to their day-to-day role in the company. The team discovered that some employees whose work focused on the "bottom of the tech stack" (e.g. on hardware or drivers) struggled to understand how "ethics" and responsible AI were relevant to their work. The Ethics & Society team helped these employees use the new core priority in their performance review as a mechanism to prioritize quality and take the time to ensure those further up the stack can use the tech responsibly. It was much easier to set core priorities and success indicators with ethical considerations for people who worked on AI that has a more directly perceivable social impact.

The Ethics & Society team also learned that managers need to have an evaluation framework so they can measure an employee's performance for promotions and bonuses. Finally, the team discovered that when upper management turned their own responsible AI core priorities from their own individual performance reviews into organization-wide Objectives and Key Results (OKRs) on responsible AI, organizational change accelerated.

This initial pilot to implement a common responsible AI core priority across Alex Kipman's organization was seen as a success within Microsoft.

Analysis of best practices

FIGURE 3

The Markkula Center for Applied Ethics has been at the forefront of applied ethics for over 30 years. Among its contributions to technology ethics are its *Ethics in Technology Practice* materials, which include an analysis of Best Ethical Practices in Technology.²⁴ In practice, Microsoft has implemented or begun to implement most of them. For example, Microsoft's Responsible AI Principles directly connect to six of the best practices.

Markkula Center for Applied Ethics best practices in technology applied by Microsoft

Microsoft	Markkula Center for Applied Ethics
Responsible AI Principles	Best practices
Fairness	aligns with
Reliability and safety	connects to
Privacy and security	matches with
Inclusiveness	connects to
Transparency	aligns with
Accountability	directly relates to

Source: World Economic Forum and Markkula Center for Applied Ethics at Santa Clara University

By turning best practices into principles, Microsoft assures that they are highly visible and likely to contribute to and influence ethical conversations.

Other Markkula Center best practices that Microsoft has incorporated include:

Keep ethics in the spotlight: The formation of the Ethics & Society team in 2017 and the expansion of its activities (e.g. the goal to create Responsible AI core priorities for every Cloud + AI team member) illustrate that Microsoft has moved beyond compliance towards inspiring culture change.

Highlight the human lives and interests behind the technology: Microsoft's user-first approach to product development demonstrates a focus on people rather than on technology. Ethical tools, such as impact assessments, Judgment Call and Community Jury institutionalize this focus as well.

Consider downstream (and upstream and lateral) risks for technologies: All of Microsoft's ethical tools are centred on considering and mitigating ethical risks. Even the more technical tools, Fairlearn, InterpretML and the error terrain analysis, serve this function by keeping an eye on the risks of bias and other problems in machine learning.

Do not discount non-technical actors, interests and expectations: Microsoft's Community Jury exercise is specially designed to bring diverse community voices into the product development effort.

Envision the technical ecosystem: The name "Ethics & Society" hints at Microsoft's recognition that it is part of a sociotechnical ecosystem. The company's willingness to share its responsible product innovation tools publicly is also a strong indication of its desire to contribute to the common good.

Treat technology as a conditional good: By choosing to develop and release some technologies but not others,²⁵ Microsoft shows that it believes technology is not an unconditional good. Not all technologies *ought* to exist. Rather, the technologies that should exist are those that help people and have positive social impact, while others should be selected against, regulated or perhaps even banned.

Make ethical reflection and practice standard, pervasive, iterative and rewarding: As Microsoft rolls out ethical practices across its organization, it is institutionalizing its ethical tools and scaling these resources to increasingly large groups. Ethical reflection is becoming more common through the Responsible AI Champs programme, company-wide education training and RAISE activation and scaling, and more rewarding through the implementation of responsible AI OKRs.

Model and advocate for ethical tech practice: By being an industry leader in integrating ethical thinking into its product life cycle and through leadership in supporting organizations, Microsoft has acted to model and advocate for ethical practices in technology. Its willingness to share some of the tools and practices is also indicative of this best practice.

What remains to be done

Microsoft has made a commitment to responsible innovation, ethics and trustworthiness. And yet, while the scope of this commitment could be limited solely to internal improvements, Microsoft has chosen to do more. Employees of the company's various ethics functions consistently state their hope that the entire technology industry will advance its ethical commitment. In this way, the company acknowledges that *ethics is not a zero-sum game*. Ethics seek to make the world a better place for everyone, and the more people and organizations strive to become and act ethically, the better it is for everyone. Microsoft's commitment sets an example for the technology industry as a whole – that technology should be designed, developed, deployed and used ethically. The tools it has developed and shared publicly are only the beginning for Microsoft and, it hopes, the world.



As AI continues to rapidly evolve, it's critical that we think carefully about the complex task of building and using it responsibly. This is a never-ending journey. As we develop new technology, we must preserve timeless values.

Brad Smith, President, Microsoft²⁶

Even with the steps that Microsoft has taken to operationalize ethics in recent years, it is not immune to regulatory and public scrutiny, especially as legal frameworks continue to evolve to satisfy public sentiments. For example, the concerns raised by European regulators about the privacy policy and practices in Office 365 are well documented.²⁷ In response, Microsoft updated its Online Services Terms for commercial cloud customers.²⁸ But these issues are not unique to Microsoft alone. The company has plans for future developments in the area of ethics and is

Microsoft is a partner in the World Economic Forum Responsible Use of Technology project. Its commitment is also reflected in its participation and leadership in the Partnership on AI to Benefit People and Society and the Vatican's Rome Call for AI Ethics, among other initiatives. Technology does not exist in a vacuum or apart from society. Technology always affects real people and human institutions. As human technological power grows, the impacts of these powerful technologies should be carefully considered and controlled *before* the effects occur. These technological impacts, as well as the preparations, controls and considerations designed to govern them, are inherently ethically charged. Therefore, technology companies must consider ethics as a part of their business. And because all companies are now technology companies, all companies should think more closely about how technology ethics is involved in their work.

investigating expanding ethical tools and processes to the entire organization. As with any major project, the company will proceed in phases, the specifics of which are being developed.

Many other companies in the technology industry are also pursuing efforts to institutionalize ethical thinking in the product development process. Some of these companies will be part of the World Economic Forum's series of case studies on the Responsible Use of Technology.

Conclusion

The rapid adoption of Fourth Industrial Revolution technologies is permeating every aspect of society. Simultaneously, the world is navigating through a global pandemic that is causing stress while creating opportunities for change. As global stakeholders cooperate to manage the direct consequences of the crisis, designing, developing, distributing, deploying and using technology responsibly are of paramount importance.

Making technology ethical will require efforts not only from technology companies but from many types of organizations worldwide. The World Economic Forum and the Markkula Center for Applied Ethics at Santa Clara University share Microsoft's journey in this endeavour in an effort to inspire and enable organizations with similar intentions to benefit from its experience. This case study aims to promote discussion, critiques, as well as efforts to build upon Microsoft's work. The World Economic Forum and its partners in this project hope more organizations not only operationalize ethics in their use of technology but also share their own experience with the global community.

BOX 1

Project Tokyo: Applying responsible innovation practices

In Cambridge, United Kingdom, a 12-year-old boy named Theo is sitting in a friend's kitchen wearing a modified Microsoft HoloLens headset. Theo is blind. When he turns his head to face a person in the room, the name of the person is played in his headset along with a bump sound. This is a Microsoft research prototype that uses artificial intelligence and augmented reality to assist people who are blind or have low vision, code named Project Tokyo.²⁹

In 2016 a team from Microsoft Research, led by Cecily Morrison, set out to explore how AI that uses computer vision could extend people's capabilities rather than replace them. As early adopters of AI technologies, the team worked alongside people who are blind or have low vision to innovate. According to Morrison, "our team wanted to imagine a future with AI that enabled people to extend their own capabilities, using a human-centric approach to developing algorithms and interactive AI experiences. A human-centric approach must also be a process of responsible innovation, a key pillar of our thinking as we developed the research."

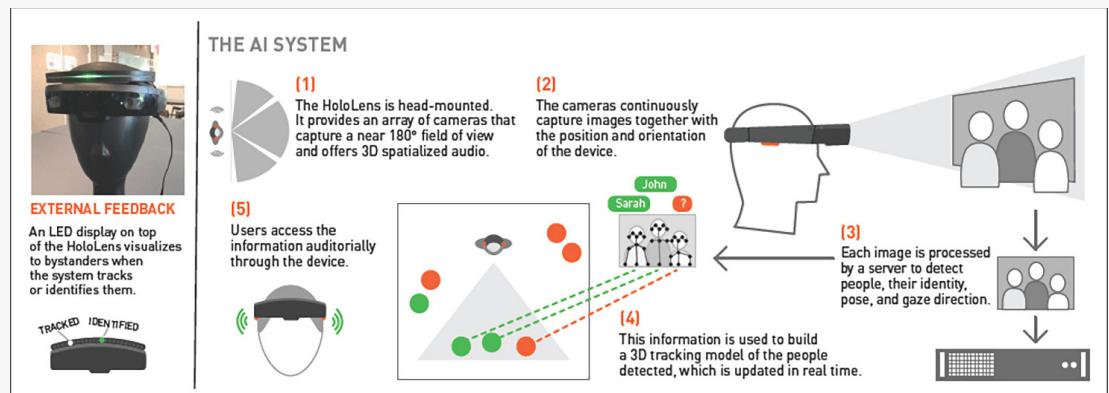
Morrison and her team of researchers began Project Tokyo by observing athletes and spectators on their trip to the Paralympic games in Brazil. From observing those who push the boundaries of

what is possible, the team learned that people who are blind or have low vision are extremely skilled at using their senses to determine what is happening around them. This is called sense-making. Yet, with less redundant information, there is a great deal of uncertainty about whether their judgements are accurate. That is where technology can come in. By providing additional information, it can help blind people and those with low vision feel more confident in their nuanced skills about making sense of the world in their own way.

The research team discovered that social information was the hardest for their participants to gather and the most important to them. A research project led the team to focus on enabling the social agency of blind children in schools, helping them understand users in their immediate vicinity.

The AI system Microsoft Research developed uses a modified HoloLens worn on the user's head to scan a 160-degree field of view. A phone or cloud server processes images from the sensors to detect people's position, gaze, pose and identity. It then communicates this information in spatial audio to the user, with sound coming from the direction of the person being identified.³⁰

On the left: Image of the adapted HoloLens device; On the right: Schematic description of the core functionality of the AI system



Source: Interactions, "Interpretability as a Dynamic of Human-AI Interaction"³¹

Sources: Interview with Cecily Morrison, Senior Researcher, Microsoft Research Lab, Cambridge, United Kingdom; Roach, John, "Using AI, people who are blind are able to find familiar faces in a room", Microsoft, Innovation Stories, 28 September 2020.

Microsoft researchers made several ethical decisions when developing the HoloLens system. Understanding the tension between inclusiveness and privacy, Morrison and her team decided to discard image, timestamp and location data to avoid misuse by a malicious actor. But that alone was not enough to make social interactions with a camera comfortable. To ensure a positive experience for bystanders identified by the system, the research team added a LED light strip to the top of the HoloLens. A white light tracks the movement of the nearest bystander. Once the system identifies that person to the user, the LED flashes green. When testing the device with Theo and his schoolmates, the children understood this interaction immediately without explanation, and began to play hide-and-seek. The research team

had enabled people to spontaneously socially negotiate the use of a technology in the moment, avoiding the need for regulation in this context.

Working through the responsible innovation process, the research team changed their mindset from thinking about the user to thinking about interactions between users – a fundamental shift in perspective that can enable the innovation of products that are thoughtful to the wider social context in which they will exist and influence.

The Office of Responsible AI at Microsoft is now using Project Tokyo as a case study to challenge its product teams to think about how engineering decisions can impact people and society in ways they may not have previously considered.

Contributors

Lead contributors

Brian Green

Director, Technology Ethics, Markkula Center for Applied Ethics, Santa Clara University, USA

Daniel Lim

Project Fellow, Artificial Intelligence and Machine Learning, World Economic Forum LLC; Seconded from Salesforce

Emily Ratté

Project Coordinator, Artificial Intelligence and Machine Learning, World Economic Forum LLC

Acknowledgements

Kathy Baxter

Principal Architect, Ethical AI Practice, Salesforce, USA

Kay Firth-Butterfield

Head, Artificial Intelligence and Machine Learning; Member of the Executive Committee, World Economic Forum LLC

Steven Mills

Partner and Chief AI Ethics Officer, Boston Consulting Group, USA

Ben Olsen

Lead, Responsible Innovation, Education and Activation, Facebook, USA

Kay Pang

Senior Director and Associate General Counsel, Global Markets Compliance Officer, VMware, USA

Ann Skeet

Senior Director, Leadership Ethics, Markkula Center for Applied Ethics, Santa Clara University, USA

Leila Topic

Head, Emerging Technologies Initiative, NetHope, USA

Thor Wasbotten

Managing Director, Markkula Center for Applied Ethics, Santa Clara University, USA

Endnotes

1. Schwab, Klaus, "The Fourth Industrial Revolution: What it means and how to respond", World Economic Forum, Agenda, 14 January 2016, <https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond> (accessed 10 February 2021).
2. Jobin, Anna, Marcello Ienca and Effy Vayena, "The global landscape of AI ethics guidelines", *Nature Machine Intelligence*, vol. 1, no. 9, 2 September 2019, pp. 389-399.
3. Skeet, Ann, and Jim Guszcza, "How businesses can create an ethical culture in the age of tech", World Economic Forum, Agenda, 7 January 2020, <https://www.weforum.org/agenda/2020/01/how-businesses-can-create-an-ethical-culture-in-the-age-of-tech> (accessed 10 February 2021).
4. Dweck, Carol, *Mindset: The New Psychology of Success*, Random House, 2006.
5. Lee, Peter, "Learning from Tay's introduction", Official Microsoft Blog, 25 March 2016, <https://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction> (accessed 11 February 2021).
6. Nadella, Satya, "The Partnership of the Future: Microsoft's CEO explores how humans and A.I. can work together to solve society's greatest challenges", *Slate*, 28 June 2016, <https://slate.com/technology/2016/06/microsoft-ceo-satya-nadella-humans-and-a-i-can-work-together-to-solve-societys-challenges.html> (accessed 11 February 2021).
7. Nadella, Satya, "Build 2017" keynote address, Microsoft Build Conference 2017, Seattle, Washington, 10 May 2017.
8. O'Brien, Tim, Steve Sweetman, Natasha Crampton and Venky Veeraraghavan, "How global tech companies can champion ethical AI", World Economic Forum, Agenda, 14 January 2020, <https://www.weforum.org/agenda/2020/01/tech-companies-ethics-responsible-ai-microsoft> (accessed 11 February 2021).
9. Lane, Mira, "Responsible Innovation: The Next Wave of Design Thinking", Medium, Microsoft Design, 19 May 2020, <https://medium.com/microsoft-design/responsible-innovation-the-next-wave-of-design-thinking-86bc9e9a8ae8> (accessed 11 February 2021).
10. World Economic Forum, in collaboration with Deloitte and the Markkula Center for Applied Ethics at Santa Clara University, "Ethics by Design: An organizational approach to responsible use of technology", White Paper, December 2020, http://www3.weforum.org/docs/WEF_Ethics_by_Design_2020.pdf (accessed 11 February 2021).
11. Nadella, "The Partnership of the Future", op. cit.
12. Microsoft, "Responsible AI", including videos, 2020, <https://www.microsoft.com/en-us/ai/responsible-ai?activetab=pivot1:primary&6> (accessed 11 February 2021).
13. World Economic Forum, Deloitte, Markkula Center for Applied Ethics, "Ethics by Design", op. cit.
14. Boyd, Eric, "Build powerful and responsible AI solutions with Azure", Microsoft, Azure Blog, 22 September 2020, <https://azure.microsoft.com/en-us/blog/build-powerful-and-responsible-ai-solutions-with-azure>; Roach, John, "Microsoft gives users control over their voice clips", Microsoft, AI Blog, 15 January 2021, <https://blogs.microsoft.com/ai/microsoft-gives-users-control-over-their-voice-clips>; Culler, Leah, "Are you talking to me? Azure AI brings iconic characters to life with Custom Neural Voice", Microsoft, AI for Business, 3 February 2021, <https://blogs.microsoft.com/ai-for-business/custom-neural-voice-ga> (all accessed 11 February 2021).
15. Ballard, Stephanie, Karen Chappell, Mira Lane and Oscar Murillo, "Judgment Call", Microsoft, 12 May 2020, <https://docs.microsoft.com/en-us/azure/architecture/guide/responsible-innovation/judgmentcall> (accessed 11 February 2021).
16. Noah, Ben, Lori Li and Arathi Sethumadhavan, "Community Jury", Microsoft, 18 May 2020, <https://docs.microsoft.com/en-us/azure/architecture/guide/responsible-innovation/community-jury> (accessed 12 February 2021).
17. Bird, Sarah, et al., "Fairlearn: A toolkit for assessing and improving fairness in AI", Microsoft, May 2020, https://www.microsoft.com/en-us/research/uploads/prod/2020/05/Fairlearn_WhitePaper-2020-09-22.pdf (accessed 12 February 2021).
18. See Fairlearn [website], "Quickstart", <https://fairlearn.github.io/v0.5.0/quickstart.html> (accessed 12 February 2021).
19. Bird, et al., "Fairlearn", op. cit. pp. 1-2.
20. Ibid., p. 6.
21. Microsoft, "EY Platform uses Fairlearn to help customers gain trust in AI", 15 May 2020, <https://customers.microsoft.com/en-us/story/809460-ey-partner-professional-services-azure-machine-learning-fairlearn> (accessed 12 February 2021).
22. Sameki, Mehrnoosh, et al., "InterpretML: A toolkit for understanding machine learning models", Microsoft, 18 May 2020, <https://www.microsoft.com/en-us/research/uploads/prod/2020/05/InterpretML-Whitepaper.pdf>; for additional resources, see <https://interpret.ml> and "interpretml/interpret" on GitHub, <https://github.com/interpretml/interpret> (all accessed 12 February 2021).
23. Barraza, Rick, et al., "Error terrain analysis for machine learning: Tools and visualizations", International Conference on Learning Representations (ICLR), 6 May 2019, <https://slideslive.com/38915701/error-terrain-analysis-for-machine-learning-tool-and-visualizations?ref=speaker-16330-latest>; Nushi, Besmira, Ece Kamar and Eric Horvitz, "Towards Accountable AI: Hybrid Human-Machine Analyses for Characterizing System Failure", Association for the Advancement of Artificial Intelligence, 2018, https://besmiranushi.com/docs/accountable_AI_hcomp_2018.pdf (both accessed 12 February 2021).

24. Vallor, Shannon, and Brian Green, "Best Ethical Practices in Technology", Markkula Center for Applied Ethics at Santa Clara University, 2018, <https://www.scu.edu/ethics-in-technology-practice/best-ethical-practices-in-technology> (accessed 12 February 2021).
25. Greene, Jay, "Microsoft won't sell police its facial-recognition technology, following similar moves by Amazon and IBM", *The Washington Post*, 11 June 2020, www.washingtonpost.com/technology/2020/06/11/microsoft-facial-recognition (accessed 12 February 2021).
26. Quote provided to the World Economic Forum for this paper.
27. Tung, Liam, "Microsoft warned: Your cloud contracts are 'serious concern' says EU privacy watchdog", ZDNet, 22 October 2019, <https://www.zdnet.com/article/microsoft-warned-your-cloud-contracts-are-serious-concern-says-eu-privacy-watchdog> (accessed 12 February 2021).
28. Brill, Julie, "Introducing more privacy transparency for our commercial cloud customers", Microsoft, EU Policy Blog, 18 November 2019, <https://blogs.microsoft.com/eupolicy/2019/11/18/introducing-privacy-transparency-commercial-cloud-customers> (accessed 12 February 2021).
29. Roach, John, "Using AI, people who are blind are able to find familiar faces in a room", Microsoft, Innovation Stories, 28 September 2020, <https://news.microsoft.com/innovation-stories/project-tokyo> (accessed 12 February 2021).
30. Thieme, Anja, et al. "Interpretability as a Dynamic of Human-AI Interaction." *Interactions*, vol. 27, no. 5, October 2020, pp. 40-45.
31. Ibid., Figure 1, 2020, <https://interactions.acm.org/archive/view/september-october-2020/interpretability-as-a-dynamic-of-human-ai-interaction> (accessed 15 February 2021).



COMMITTED TO
IMPROVING THE STATE
OF THE WORLD

The World Economic Forum, committed to improving the state of the world, is the International Organization for Public-Private Cooperation.

The Forum engages the foremost political, business and other leaders of society to shape global, regional and industry agendas.

World Economic Forum
91–93 route de la Capite
CH-1223 Cologny/Geneva
Switzerland

Tel.: +41 (0) 22 869 1212
Fax: +41 (0) 22 786 2744
contact@weforum.org
www.weforum.org