

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

WORLD  
ECONOMIC  
FORUM

# Responsible Use of Technology: The IBM Case Study

WHITE PAPER  
SEPTEMBER 2021



# Contents

3	Foreword
4	Introduction
5	1 The evolution of artificial intelligence ethics at IBM
8	2 Governance
8	2.1 AI Ethics Board
8	2.2 Local focal points and Advocacy Network
9	3 IBM's Principles for Trust and Transparency
9	3.1 The purpose of AI is to augment human intelligence
9	3.2 Data and insights belong to their creator
9	3.3 New technology, including AI systems, must be transparent and explainable
10	4 The pillars of trust
10	4.1 Explainability
10	4.2 Fairness
10	4.3 Robustness
10	4.4 Transparency
10	4.5 Privacy
11	5 Trustworthy AI research and toolkits
11	5.1 Open source
14	5.2 IBM proprietary tools
15	6 Delivery capabilities for ethical outcomes
15	6.1 Trustworthy AI
16	6.2 Ethics by design
17	7 Leveraging AI in the workplace
18	8 Diversity and inclusion
19	9 Education and enablement
20	10 Partnerships and stakeholder engagements
22	11 Summary of progress
22	11.1 Influencing policy
22	11.2 Integrating ethics in the AI development pipeline
23	12 What remains to be done
24	Conclusion
25	Contributors
26	Endnotes

## Disclaimer

This document is published by the World Economic Forum as a contribution to a project, insight area or interaction. The findings, interpretations and conclusions expressed herein are a result of a collaborative process facilitated and endorsed by the World Economic Forum but whose results do not necessarily represent the views of the World Economic Forum, nor the entirety of its Members, Partners or other stakeholders.

© 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

Advanced technologies, such as artificial intelligence (AI), distributed ledger and cloud computing, are at the heart of the Fourth Industrial Revolution. These technologies, deployed appropriately, can reinvent government services, education, healthcare, and the way businesses interact with their customers. In the process, they can also unlock moral value and benefits to individuals and society. If not responsibly directed, however, Fourth Industrial Revolution technologies have the potential to erode trust, exacerbate inequity, and cause harm to people and the environment.

While the COVID-19 pandemic has brought uncertainty to all facets of society, one thing is certain: the importance of technology in daily life has never been greater. In the past year alone, AI has been used to accelerate the pace of vaccine research, the application of distributed ledgers to securely deliver digital health passports, and the availability of speech-to-text transcription in every online classroom. With so much at stake, it is essential that we, as a society, continue to advance the practice of infusing ethics and responsibility through the entire life cycle of technology. To this end, the World Economic Forum Centre for the Fourth Industrial Revolution has been leading the [Responsible Use of Technology project](#), which brings together a community of experts from civil society, governments and companies to provide practical tools for leaders 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. Since the inception of this project, we established the need for both the human rights-based and ethics-based approaches to the [responsible use of technology](#). We adapted concepts of behavioural economics to help leaders [design organizations](#) to drive more ethical outcomes with technology. We also convened numerous workshops and [public discussions](#), during which practitioners shared their experiences in applying ethics to technology.

To continue the knowledge sharing and advance the practice of responsible technology, the World Economic Forum and the Markkula Center for Applied Ethics at Santa Clara University are collaborating to publish a series of case studies highlighting processes, tools and organizational constructs that facilitate the responsible design, development and implementation of technology. This document, “Responsible Use of Technology: The IBM Case Study”, marks the second in a series, following a White Paper on [Microsoft](#). We would like to thank IBM for sharing their ethical technology governance structure, practices, tools, activities and research expertise for this effort. It is our hope that this document and the Responsible Use of Technology project will enable other organizations to learn from the insights of our community and take steps to incorporate ethics into their design, development, adoption and use of technology.

# Introduction

Technological breakthroughs are permeating society and creating new opportunities and challenges every day. For those who are subject to these changes, as well as those who develop new technologies, the social impact can be concerning.

In response to these concerns – not only in society, but also among the researchers and developers of these technologies – many companies have sought ways to ensure their technologies are designed, developed and used responsibly. This White Paper focuses on IBM, a company that has long-standing commitments to ethics and privacy, and a multistakeholder approach to encourage beneficial interactions between technology and

society. Its research division has been working for years towards developing tools and processes for responsible technology. The company has worked to develop a culture which thinks of the ethical impact of its technology, and how to steer new technologies towards their better uses.

Through a series of interviews with IBM executives and employees, combined with secondary research, this paper presents an external view of IBM's experience in transforming the company towards thinking more deeply about technology ethics and how to work to responsibly design, develop, deploy and use technology.

## BOX 1 The responsible use of technology initiative

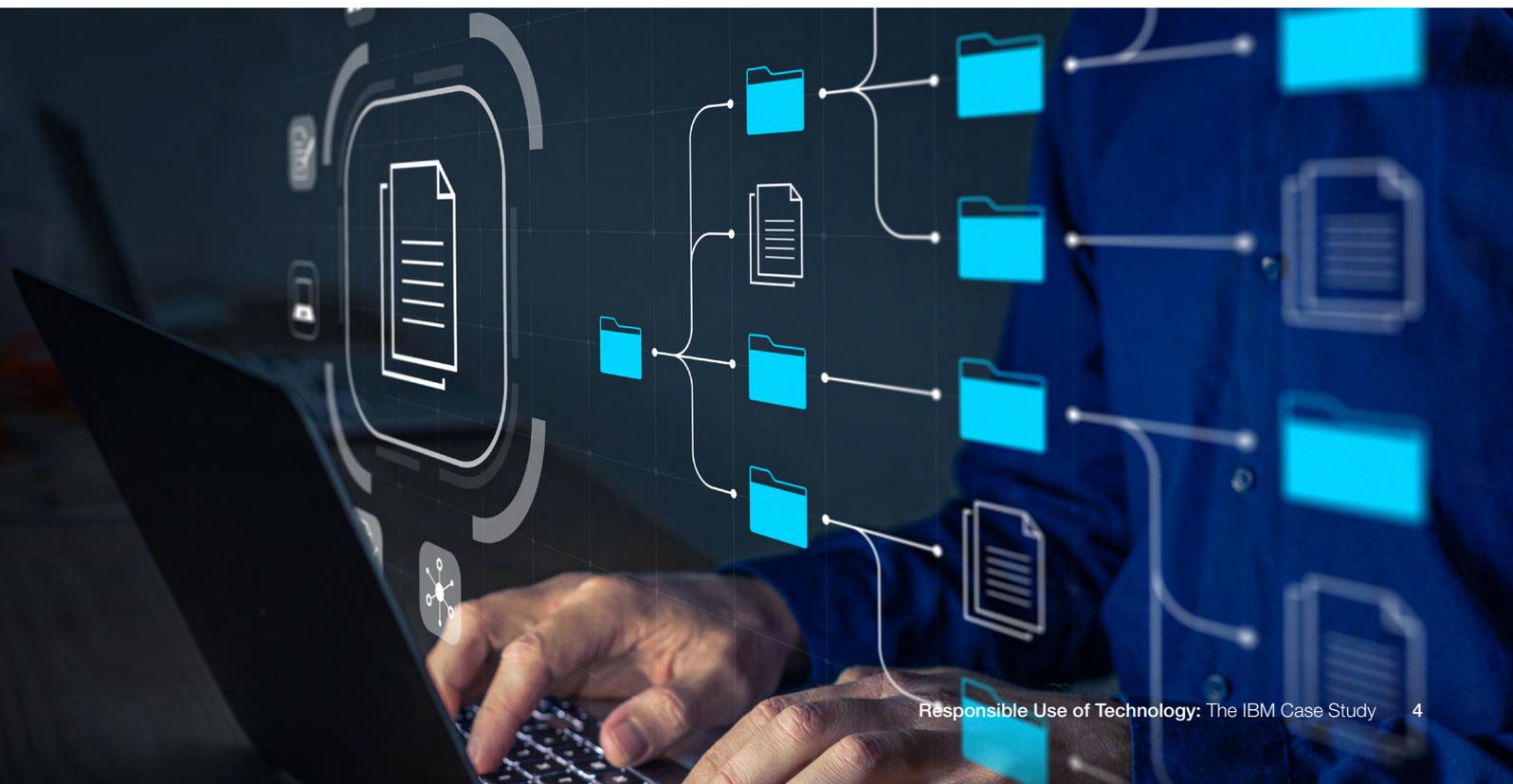
Perhaps nothing, not even the race to the moon, has changed society with such speed and scale as the recent advances of new, disruptive technologies.

Yet as much as technological innovation is revered, speed also needs to be managed. Many companies and countries are realizing that the consequences of Fourth Industrial Revolution technology go hand in hand with the new playing fields it creates for developing and selling beneficial products. They are now confronting growing “techlash” with more people demanding greater accountability from tech companies than ever before.

Self-regulation will only go so far. An urgent need exists for practical guides, based on proven

models, that companies everywhere can use to ensure that ethical considerations are integrated into company processes and cultures, including in product design and development.

The World Economic Forum Responsible Use of Technology initiative aims to provide useful resources for organizations to operationalize ethics in technology. Sharing experiences and best practices among “ethics executives”, as well as developing tools and frameworks through consultations with members of business, government and civil society, is thus part of this initiative of the World Economic Forum Platform for Shaping the Future of Technology Governance: Artificial Intelligence and Machine Learning.



1

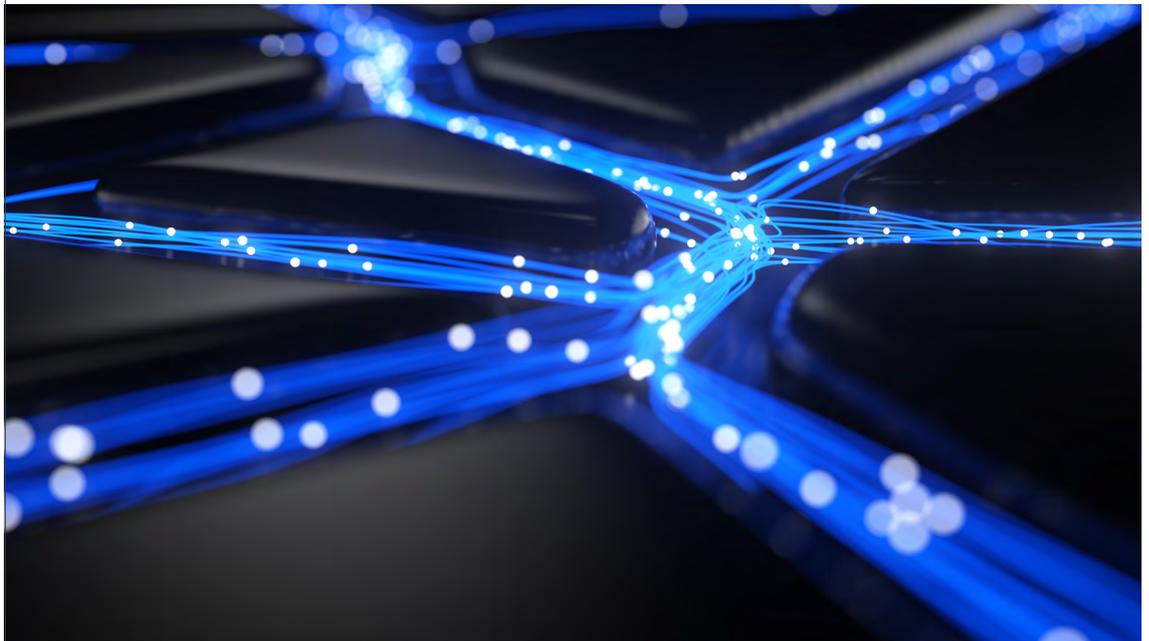
# The evolution of artificial intelligence ethics at IBM

As the oldest major technology company, IBM has built its corporate culture for more than a century. As artificial intelligence (AI) ethics was becoming an increasingly important topic in technology, IBM was in many ways prepared for these discussions thanks to the ethical interests of its employees and to its corporate culture, which encouraged those interests.

To get different parts of the company working together, IBM embarked on a journey in 2003 to redefine its company values. A three-day online discussion with employees led to the following three values: 1) dedication to every client's success; 2) innovation that matters, for the company and for the world; and 3) trust and personal responsibility in all relationships.<sup>1</sup> The

three values stand on their own but are also intrinsically linked; IBM values its clients and wants them to succeed. Within this relationship, trust makes success possible – not only trust in people, but also trust in the technology that people have built. Working with businesses, such as banks and hospitals, as well as with governments in long-term relationships, means that trust is a paramount concern for IBM. Business relationships are commitments to be protected and fulfilled.

In addition to its relationships with clients, IBM recognizes that it exists in a social context along with governments and the public. Thus, it looks to make itself trustworthy to these groups and, indeed, to all stakeholders.

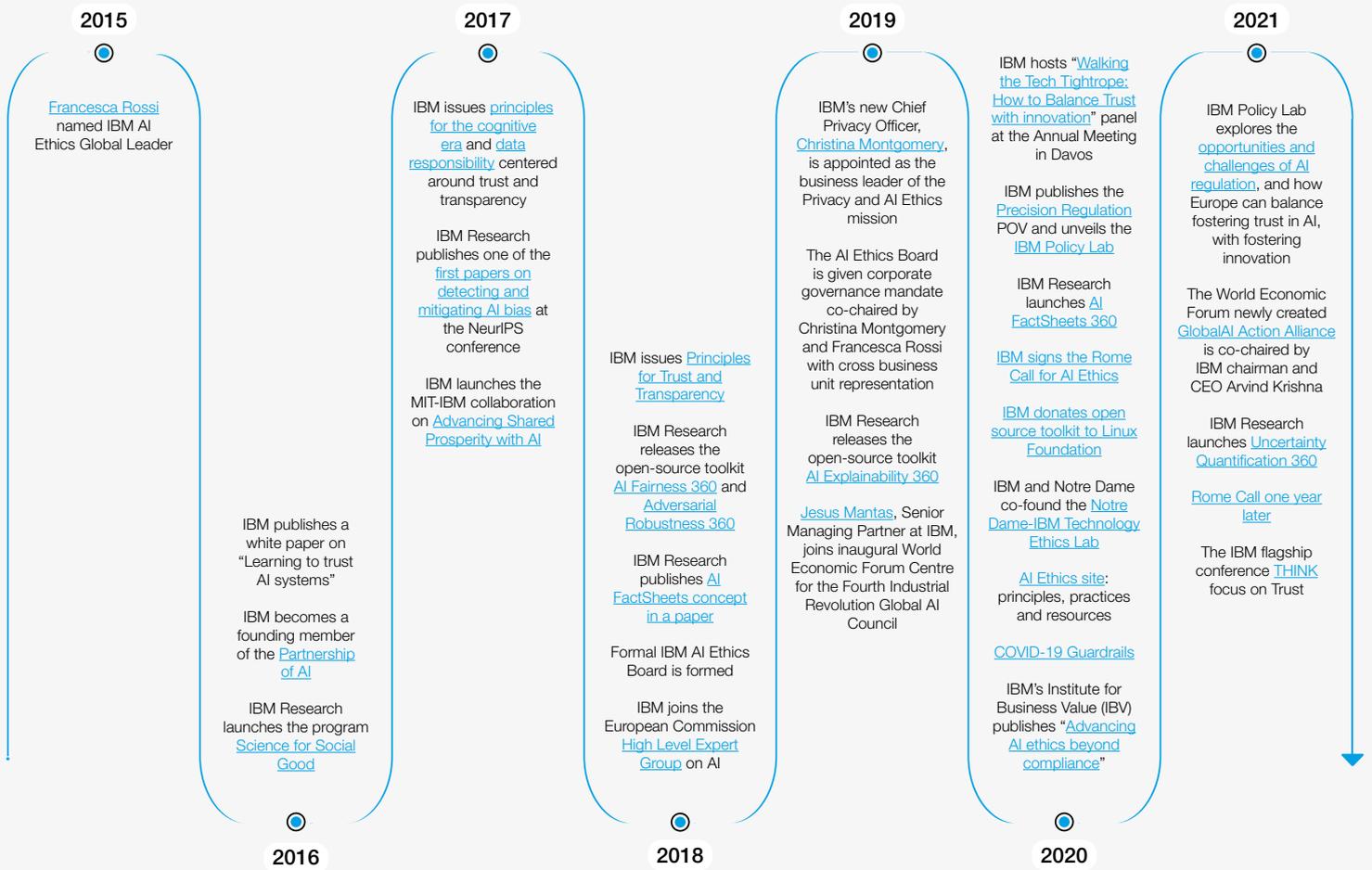


“

**Going forward, we are going to compete on trust. On values. And on our responsibility to the wider community of people whose lives we touch.**

Arvind Krishna, Chairman and Chief Executive Officer, IBM, USA<sup>2</sup>

FIGURE 1 | Responsible and ethical development of AI technology at IBM, 2015-2021



Source: Image provided by IBM

Several key moments in the past years (Figure 1) have helped propel IBM on its journey towards operationalizing responsible and ethical development of AI technology.

Francesca Rossi, Global Leader of AI Ethics at IBM, joined the company in 2015 with a mandate to explore the topic of AI ethics. She convened a group of about 40 colleagues at the company who soon published a White Paper<sup>3</sup> that listed five commitments IBM planned to make to advance its understanding and to effect the ethical development of AI. They are:

- The establishment of an internal IBM AI Ethics Board to discuss, advise and guide (eventually govern) the ethical development and deployment of AI systems (by IBM and its clients)
- A company-wide educational curriculum on the ethical development of AI
- The creation of the IBM Artificial Intelligence, Ethics and Society programme, a multidisciplinary research programme for the

ongoing exploration of responsible development of AI systems aligned with the organization's values

- Participation in cross-industry, government and scientific initiatives and events on AI and ethics
- Regular, ongoing IBM-hosted engagement with a robust ecosystem of academics, researchers, policy-makers, non-governmental organizations (NGOs) and business leaders on the ethical implications of AI

This paper presents how IBM is following through on these commitments, from principles to implementation, and from governance to technological products.

Companies developing new technologies always encounter challenges along the way. For instance, the 2018 Gender Shades study found that IBM's gender classification in facial recognition software, along with those of several other organizations, had a disproportionately high error rate when looking at darker-skinned females.<sup>4</sup> In response, IBM upgraded its classification model.<sup>5</sup> IBM Watson Health also

received negative press in 2018 about the training performance of its diagnosis product. The media coverage raised questions about the expectations of Watson Health for oncology. These stumbles are a good reminder that ethics is an ongoing process requiring continual efforts, and that remaining humble is itself important for ethical progress.

The 2018 Gender Shades study is an example of the kind of opportunity labelled a “turning point”, as it typically is in response to an external event or crisis. IBM’s choice to respond to this study reflects its sense of responsibility to society, a fundamental cultural element and one of the key conditions in companies that use ethics regularly and easily.<sup>6</sup>

IBM’s history reveals an organizational culture that makes an effort to reflect and learn from experiences and to encourage interconnectedness between people. It practises organizational introspection by reviewing its history, assessing its present position and acknowledging uncertainty.<sup>7</sup> While IBM does not claim to have all the answers, it hopes those it has found, and the lessons it has learned, might be useful to others who seek to understand more about operationalizing ethical thinking in tech companies and their own organizations.



2

# Governance

## 2.1 AI Ethics Board

At the centre of IBM’s trustworthy AI efforts is the internal AI Ethics Board (Figure 2). The concept of this Board began taking shape in 2016 when a working group of IBM employees committed to forming a central body to discuss, advise on and guide the ethical development and deployment of AI at the company. This group led to the first formal version of the Board in 2018, which guided the creation of IBM’s Principles for Trust and Transparency and the Pillars of Trust. The values outlined in the Principles and Pillars set the foundation for IBM’s AI ethics efforts in the future.

As the AI Ethics Board began to consider the future of the company’s technology ethics, as well as concrete actions to achieve them, the need for more decision-making power and reach across the company became apparent. With this in mind, IBM reorganized the Board in 2019 to be co-chaired by Christina Montgomery, IBM Chief Privacy Officer, and Francesca Rossi, Global Leader, AI Ethics, with representation from senior leaders from its business

units and corporate functions (e.g. research, professional services, human resources, diversity and inclusion, legal, government and regulatory affairs, procurement, communications). With this reorganization, the Board now has the authority to implement, enable and enforce its decisions. The Board receives its oversight from the Privacy Advisory Committee, which includes privacy and AI Ethics. The Board and the dedicated Project Office oversees the management of technology ethics initiatives and teams across IBM’s business units. For example, the Board guided and supported IBM’s decision to stop offering, researching or developing general purpose facial recognition or analysis software.<sup>8</sup>

More recently, the Board has been working on employee education on AI ethics. Importantly, the Board is not just focused on AI; it is responsible for providing governance for the ethical issues pertaining to technology overall at the company, such as supporting IBM’s response to the COVID-19 pandemic.

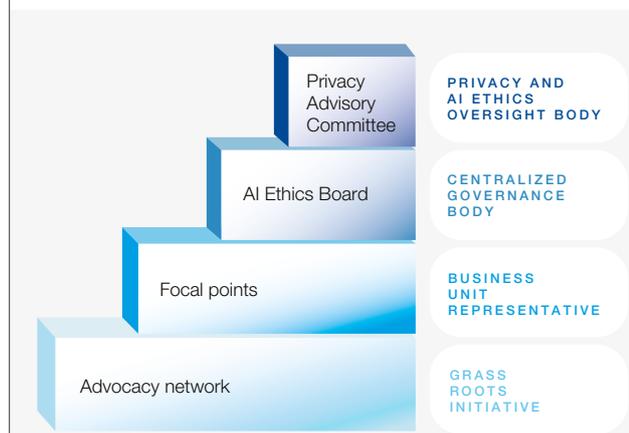
## 2.2 Local focal points and Advocacy Network

In addition to the diverse representation of departments on the AI Ethics Board, each IBM business unit has “focal points”, or individuals whose role is to support the Board’s work by engaging directly with their business units to address questions, provide insights as well as gather feedback so that the Board’s programmes can continuously improve. The focal points also work with their business units to assess whether their products and services adhere to the company’s Principles for Trust and Transparency, described below. When the local business units and the focal points have questions or concerns during their ethical assessments, the Board offers advice or determines the appropriate course of action.

IBM also has an Advocacy Network of ethical technology champions throughout the company. This network of volunteers helps to promote a culture of ethical, responsible and trustworthy technology. Occasionally, these volunteers contact the AI Ethics Board to showcase the grassroots ethics initiatives of the business units. For example, a volunteer from the network reached out to the Project Office to present an AI Ethics chatbot

to the rest of IBM that was created for a client. The business unit, the volunteer and the Project Office partnered to adapt the chatbot (now called “Trustbot”) for internal use.

FIGURE 2 | Governance structure for IBM’s AI efforts



Source: Graphic provided by IBM

3

# IBM's Principles for Trust and Transparency

The above governance efforts involve procedural and structural changes at IBM based on its company values. As a further expression of these values, IBM's principles, pillars and toolkits work to realize those values in a measurable, replicable way in a product setting.

To begin this actualization, IBM first chose three Principles for Trust and Transparency, published in

2018.<sup>9</sup> With them, IBM seeks to earn the trust of its clients and society by being trustworthy, not only through creating products that can be understood, but also by maintaining appropriate relationships with data owners and those subject to AI, such as workers. Importantly, IBM is not just seeking to get people to "trust us", but is also looking to demonstrate why it is worthy of trust, a crucial difference sometimes overlooked in discussions of trust and AI systems.

## 3.1 The purpose of AI is to augment human intelligence

AI is sometimes perceived as an adversary of human labour, automating tasks that are taken away from people. By instead considering AI as a tool to help people fulfil their tasks, a more harmonious relationship with AI is possible.

Obviously, this is easier said than done, so further processes and tools have been, are being and will need to be developed to help fulfil this principle. Many of these tools focus on education and training.

## 3.2 Data and insights belong to their creator

With this principle, IBM keeps data ownership where it belongs – with the people and organizations from whom the data originated. Clients own the data even when it is on IBM's

servers or has been processed through IBM AI. This data ownership principle has important ramifications for privacy, security, government access and cross-border data flow.

## 3.3 New technology, including AI systems, must be transparent and explainable

With this principle, IBM seeks to maintain assurance that AI systems are working in ways that make sense to the people using them. The company should be clear about when AI is being used and

what data is involved in that AI (while maintaining client data ownerships and all that they entail). It should address bias proactively, and generally seek to make AI understandable.

# 4

# The pillars of trust

IBM's pillars of trust help to implement the Principles of Trust and Transparency.<sup>10</sup> Each pillar is a well-established research subfield with dedicated communities. IBM Research has published dozens of papers in these fields and created open-source toolkits to bridge the gap between scientists and

practitioners to enable them to build trustworthy AI models. The pillars act as mid-level principles that focus the larger picture of building AI trustworthiness into more precise applications and ensure operationalization is thorough and complete, all the way to the level of code.

## 4.1 Explainability

This pillar focuses on the question of why an AI model made a particular decision, such as denying a loan application. AI-powered decisions should be accompanied by the reasons and explanations for the decisions. Since some AI systems are not

understandable even to an expert, this task is challenging. Several approaches to this problem have been proposed and made available through the IBM AI Explainability 360 toolkit (see below).

## 4.2 Fairness

Fairness looks at the issue of treating different groups equally, such as different genders having similar loan approval rates in a training data set or

in the recommendations of an AI model. This pillar directly connects to the IBM AI Fairness 360 toolkit for detecting biases in AI (see below).

## 4.3 Robustness

AI systems must be resistant to attacks. Robust systems can cope with attacks without causing serious harm; particularly with AI, these systems can cope with adversarial attacks directed in ways

that target the weakness specific to AI systems. The IBM Adversarial Robustness 360 toolbox contains defences against such attacks (see below).

## 4.4 Transparency

AI systems must be transparent to create trustworthy AI: people need to be able to understand and evaluate how the systems work and know their strengths and weaknesses. This

requires disclosure of relevant aspects of those systems. The IBM AI FactSheets 360 tool allows AI developers to collect and document facts about their models to be fully transparent (see below).

## 4.5 Privacy

People deserve to have their sensitive data protected. This includes notifying people how and when their data is being used, and what it is being used for and by whom. Only necessary data should be collected,

and consumers should have clear access to controls over how their data is being used. IBM has a set of privacy resources to support this pillar.<sup>11</sup>

5

# Trustworthy AI research and toolkits

IBM Research Labs have helped to create innovations that are injected into products, services and toolkits for both internal and client-facing solutions. For several years, AI Research at IBM has included a large division of researchers focused on trustworthy AI in areas such as AI fairness and explainability, producing over 200 publications.<sup>12</sup> This IBM Research division supported the definition of the pillars of trustworthy AI and created the open-source toolkits supporting the pillars.

Trustworthy AI toolkits include both open-source and proprietary solutions. Every toolkit

has an extensive website describing its content and uses, and the open-source algorithms are available on the development platform GitHub. The development of open-source code also demonstrates IBM’s commitment to not only develop tools for itself, but also to help the entire industry move towards and adopt responsible and trustworthy AI. GitHub engagement statistics indicate that an active community follows and contributes to the open-source tools. Over 1,300 people have copied the code to work with it, “forking” the tools to work on them, and thousands more have noted them positively.

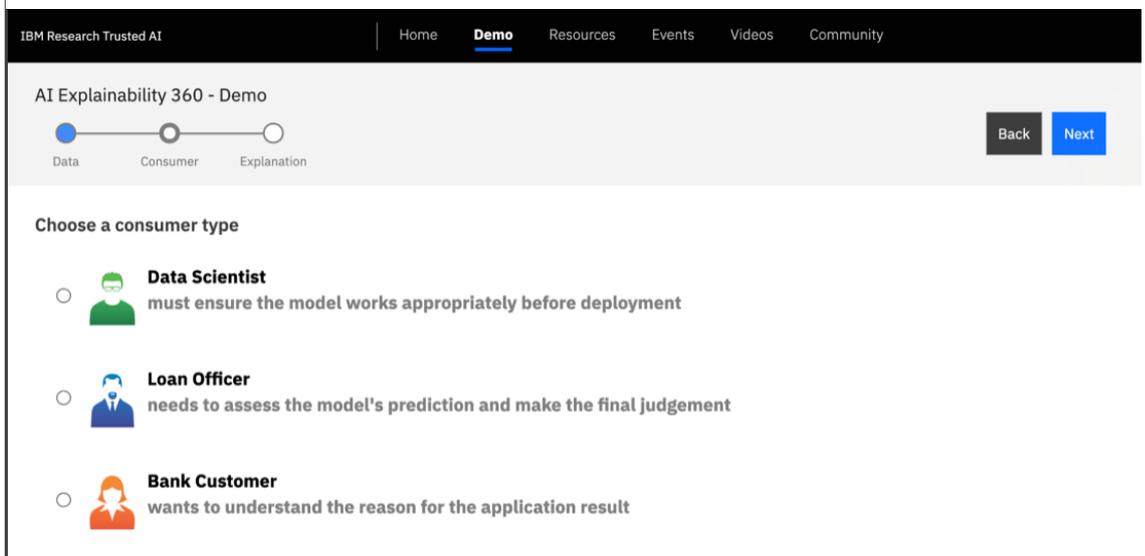
## 5.1 Open source

In 2020, IBM donated a set of AI toolkits to the Linux Foundation, including AI Explainability 360 and AI Fairness 360, as well as the Adversarial Robustness 360 toolbox. IBM has also open-sourced AI FactSheets 360 and Uncertainty Quantification 360 toolkits. These toolkits, described below, facilitate collaboration among researchers, practitioners and the multidiscipline trustworthy AI community.

The [AI Explainability 360 toolkit](#) supports IBM’s commitment to explainability, its first pillar of

trust. This toolkit has eight algorithms for making machine-learning (ML) models more transparent and comprehensible, including a demo to help understand how various users require different explanations (Figure 3). Making such models comprehensible is vital to also making them trustworthy. After all, if models work most of the time, but occasionally fail inexplicably, people are unlikely to place their trust in the model, especially if those failures create a risk of harm.<sup>13</sup>

FIGURE 3 AI Explainability 360 demo

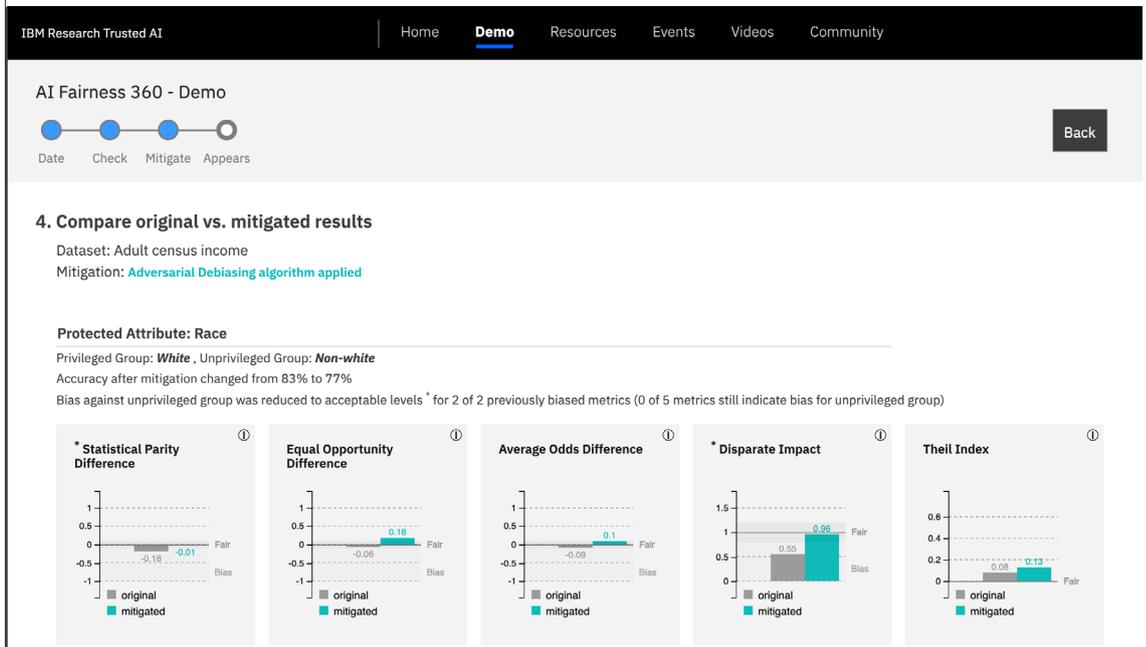


Source: IBM, “AI Explainability 360 – Demo”, <https://aix360.mybluemix.net/consumer> (accessed 16 August 2021)

IBM supports its commitment to fairness, the second pillar, through the [AI Fairness 360 toolkit](#). The toolkit includes more than 70 fairness metrics and 10 bias mitigation algorithms for enhancing fairness in ML, as well as a demo (Figure 4). In

addition, it contains metrics for assessing whether individuals and groups are treated fairly in the implementation of AI. This toolkit is based on IBM's research on AI fairness as featured in a 2018 paper.<sup>14</sup>

FIGURE 4 AI Fairness 360 demo

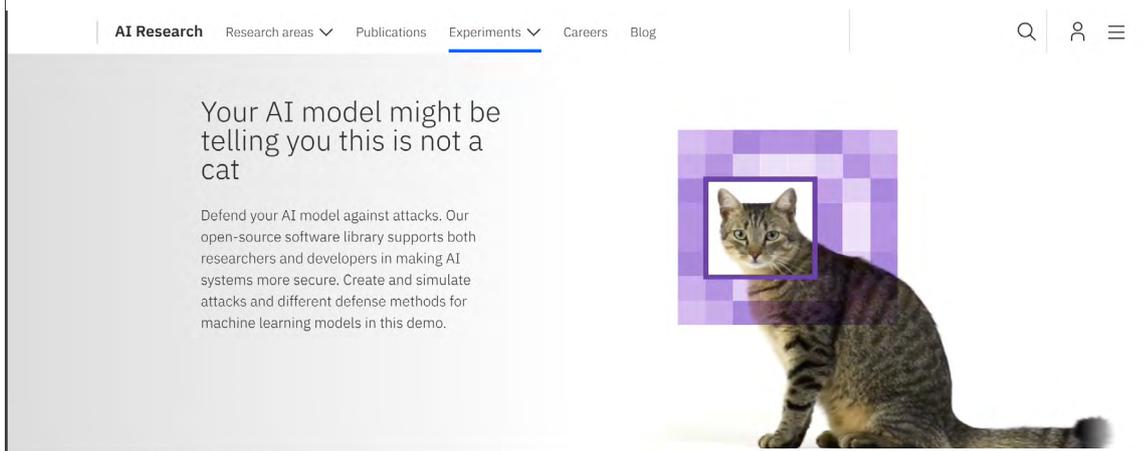


Source: IBM, from “AI Fairness 360 - Demo”, <https://aif360.mybluemix.net/data> (accessed 16 August 2021)

The [Adversarial Robustness 360 toolbox](#) supports IBM's commitment to robustness, its third pillar of trust. The tools available allow ML developers to defend against adversarial attacks, such as poisoning, inference, extraction and evasion. The

toolbox includes a demo to create and replicate attacks and defence methods for ML models (Figure 5) so that developers can evaluate how their models will do when subjected to such attacks, in order to adjust the models and make them more robust.<sup>15</sup>

FIGURE 5 Adversarial Robustness demo

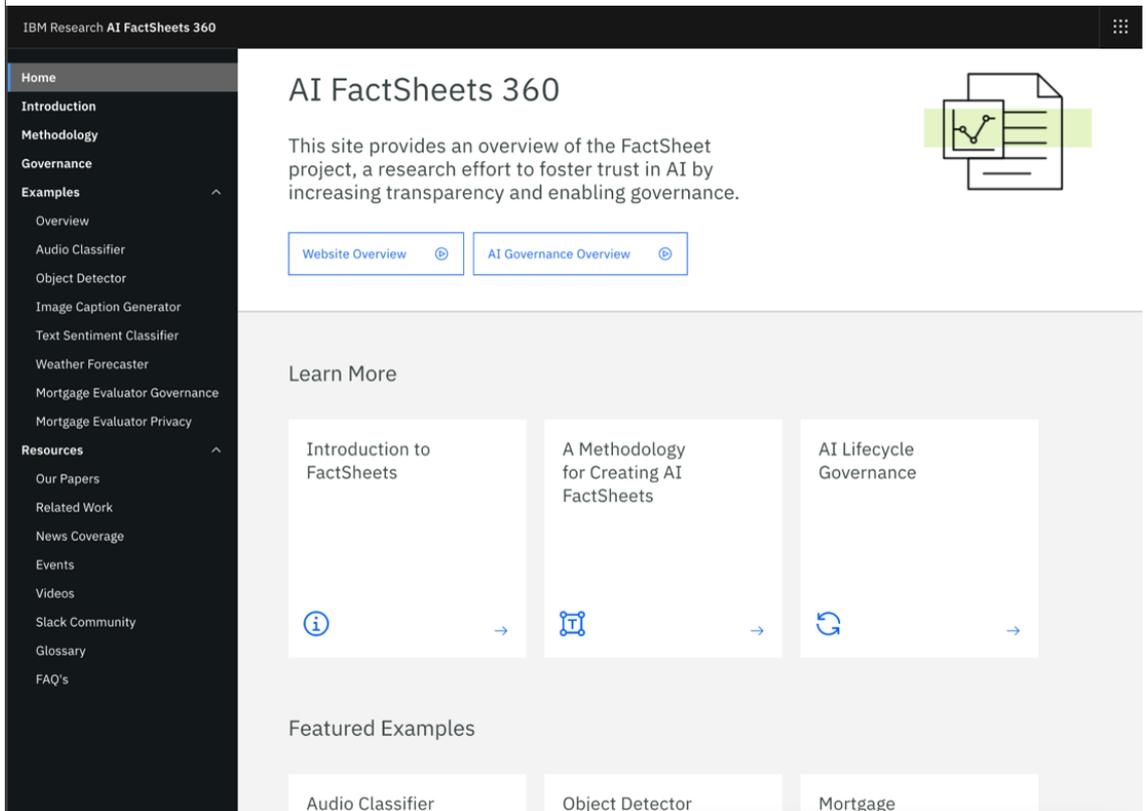


Source: IBM, “Your AI model might be telling you this is not a cat”, <https://art-demo.mybluemix.net> (accessed 16 August 2021)

The company supports its commitment to the fourth pillar of trust, transparency, through the [AI FactSheets 360 tool](#) (Figure 6). Factsheets are collections of facts about an AI model. The tool has several examples of FactSheets as well as a methodology for creating them. The information collected by these sheets is useful

for understanding the model and enabling its governance. Facts include the purpose of the model, characteristics of the data set and various relevant aspects of the creation or deployment of the model or service.<sup>16</sup> This tool is based on IBM's research efforts on AI trust, as highlighted in a seminal paper published in 2018.<sup>17</sup>

FIGURE 6 AI FactSheets 360

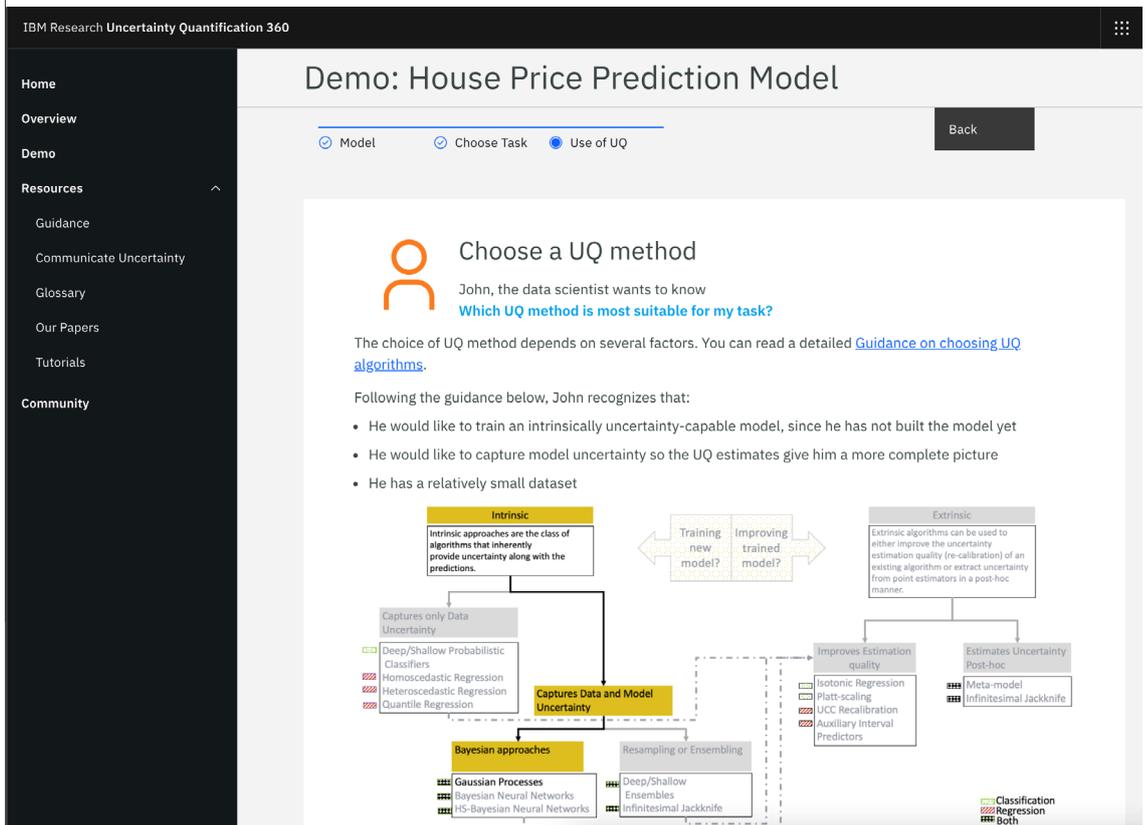


Source: IBM, "AI FactSheets 360", <https://aifs360.mybluemix.net> (accessed 16 August 2021)

The [Uncertainty Quantification 360 toolkit](#), which includes a demo (Figure 7), supports several of IBM's pillars of trust, especially robustness and explainability. By developing a better understanding of the reliability of AI judgments in particular

situations, uncertainty quantification reveals the importance of intellectual humility and helps to prevent overconfidence in AI systems by clearly indicating their capabilities and limitations.<sup>18</sup>

FIGURE 7 Uncertainty Quantification 360 demo



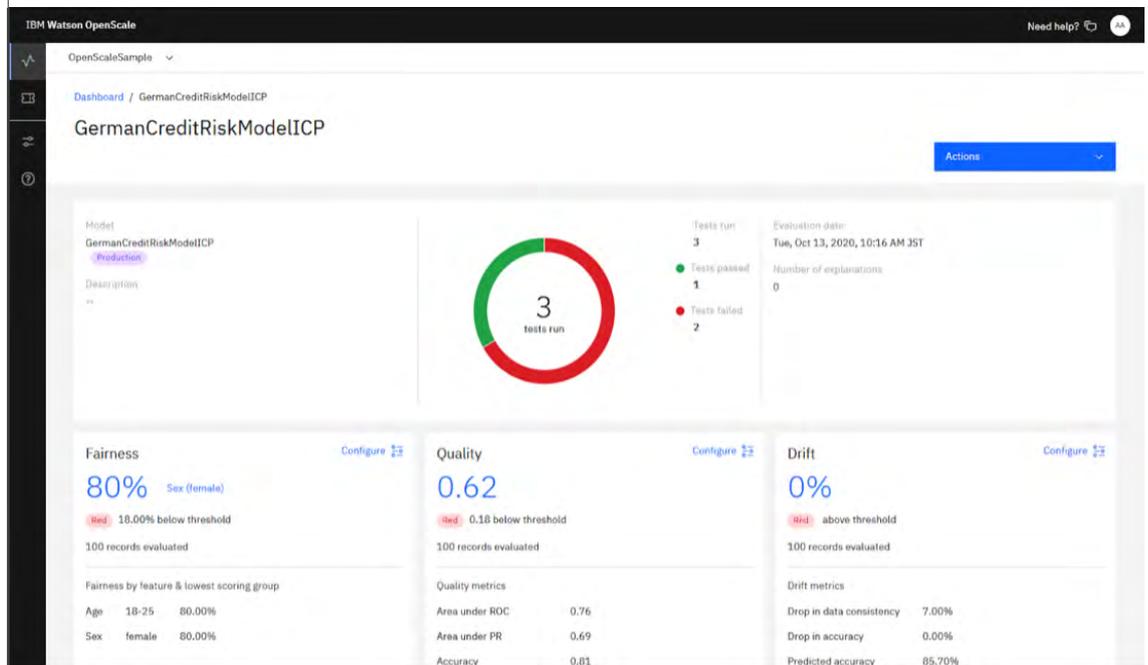
Source: IBM, "Demo: House Price Prediction Model", <http://uq360.mybluemix.net/demo/2/choose> (accessed 16 August 2021)

## 5.2 IBM proprietary tools

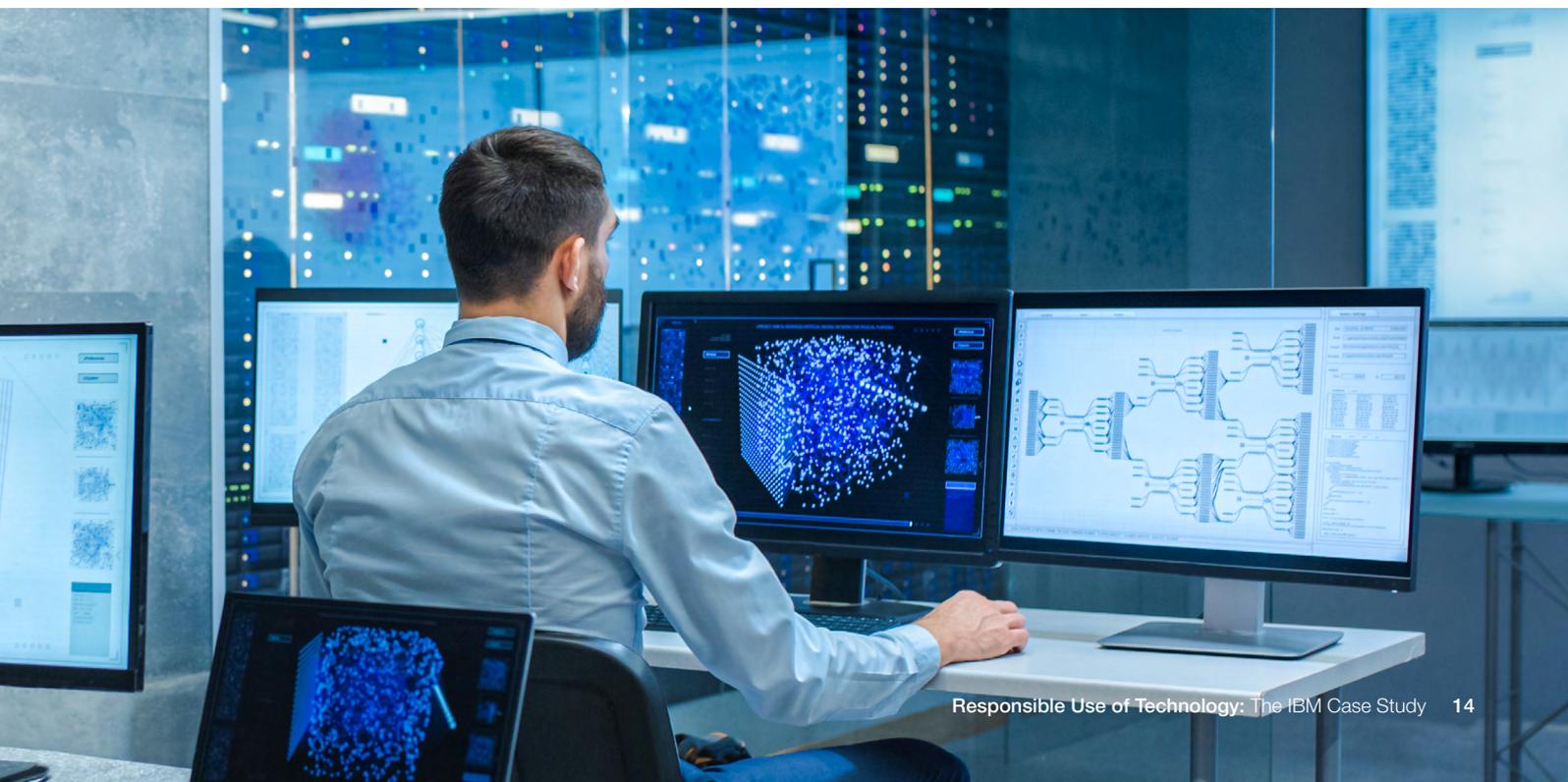
IBM Cloud Pak for Data is a consolidated set of proprietary tools for use in the entire data analytics life cycles. It allows all users from an enterprise, for example business users, data scientists and ML operations, to collaborate on the same toolset to collect, organize and analyse data. A capability in Cloud Pak for Data is Watson Studio (Figure 8), which allows users to monitor and manage AI models to operate trustworthy AI. Watson Studio monitors ML models across several data quality metrics, and on an ongoing basis for drift, which represents the degree by which ML models have moved away from the ground truth (the vetted data used to adapt [Watson Studio](#) to a particular domain<sup>19</sup>).

The tool can also monitor ML models for bias; it allows users to identify populations that are potentially at risk for bias, as well as the favourable and unfavourable outcomes in a model. As the model makes recommendations, Watson Studio can identify a bias in the recommendations against the identified target populations. Lastly, the tool presents the recommendations and areas of bias in ways that business users and data scientists can understand, and can also provide explanations.

FIGURE 8 Watson Studio



Source: IBM, "Manage model risk – Perform analysis in Watson OpenScale", <https://www.ibm.com/docs/en/cloud-paks/cp-data/3.5.0?topic=governance-manage-model-risk> (accessed 16 August 2021)



6

# Delivery capabilities for ethical outcomes

IBM believes that achieving trustworthy AI requires targeted governmental regulations and the adoption of industry best practices anchored in ethical principles. Led by the AI Ethics Board, the company

has developed capabilities such as trustworthy AI and ethics by design to help its employees and clients create and ensure more ethical outcomes.

## 6.1 Trustworthy AI

The tools described above help IBM and its clients build and adopt trustworthy AI. These tools complement the company's range of client

trustworthy AI services throughout the full AI life cycle, including the following (Figure 9):

FIGURE 9 Trustworthy AI



### IBM'S TRUSTWORTHY AI APPROACH WITH CLIENTS

IBM defines trustworthy AI as fair, explainable, transparent, privacy-preserving and robust  
Data science, design teams and industry experts  
Design thinking for data and AI  
Trustworthy AI tools  
Governance frameworks  
Open partner ecosystem

#### ASSESS, AUDIT AND MITIGATE RISK

Guidance and tooling to help clients access, audit and mitigate risk in existing AI solutions



#### FULL AI LIFE CYCLE

Partnering with clients to plan, build, deploy and manage new AI solutions while ensuring trustworthiness



#### AI GOVERNANCE FRAMEWORKS

Partnering with clients to set up an AI governance framework and implement enterprise-scale AI governance solutions



#### EDUCATION AND GUIDANCE

Best practices for building trustworthy AI solutions with education and guidance for data scientists, developers and decision-makers



Stand-alone courses and certifications or integrated with above solutions

Source: Figure provided by IBM



## 6.2 Ethics by design

IBM believes that design plays an important role in providing the guard rails on how people interact with technologies. The company regularly holds stand-alone design thinking exercises for AI ethics that focus on different topics or areas of ethical AI design, such as understanding the ripple effects of decisions or mitigating stakeholder tensions. These activities are detailed in frameworks and training, such as [team essentials for AI](#). IBM also holds larger AI ethics-focused workshops with associated sprint planning activities. With the support of the AI Ethics Board, this ongoing work is scaling across different business units.<sup>20,21</sup>

The company combines design thinking, lean start-ups, pair programming and continuous integration in IBM Garage, its iterative delivery model. From a responsible innovation perspective, the IBM Garage

methodology includes autonomous squads that co-create, co-execute and co-operate the technology solution. Specifically, the co-create phase incorporates ethical consideration exercises, such as empathy mapping, a design-thinking technique that helps squads to build a greater understanding of users and their communities.<sup>22</sup> The methodology also includes a step for “what-if” analysis,<sup>23</sup> which reminds squads to consider the potential effects of their technologies on users and communities. Research conducted at the World Economic Forum<sup>24</sup> demonstrated that organizations combining a set of strong ethical values with activities that help their teams connect with their communities have a greater likelihood of creating technology solutions that are more beneficial to society. The IBM Garage methodology is an example of such an approach.

7

# Leveraging AI in the workplace

IBM applies ethical principles to the AI tools it uses internally. The company does so based on the underlying principle that AI provides insights but does not replace human decisions. IBM trains its human resources (HR) professionals on the concept that everyone has a role to play in developing and deploying trustworthy AI.

This begins with IBM's hiring practices, where recruiters are prompted to ensure hiring managers are presented with diverse candidates. In this setting, IBM uses a suite of assessment tools that help measure technical skills, cognitive ability or learning agility. These assessments provide managers with evidence-based data points they can use to assist them in hiring decisions. Using post-hire performance data, these assessments have been validated internally, and adverse impact analysis is conducted to help ensure equal opportunity. As a result, the proportion of under-represented minorities among new hires increased from 18.9% in 2018 to 21.3% in 2020.<sup>25</sup>

IBM also uses AI to assist with personalized salary increase recommendations. Fairness assessments identify and mitigate potential bias

in the recommendations. Managers receive salary range recommendations for each employee, along with the reasons for such recommendations. Managers exercise (or retain) discretion when acting on those recommendations, and employees are told how they are paid relative to the market. The AI recommendations give data points for managers to consider, but the decision-making and accountability remains with people. IBM found that attrition is reduced by one-third when managers follow the AI recommendations. IBM's HR department is developing factsheets for employees to help them understand how AI is being used in the skills management system.

To promote open discussion regarding HR's use of AI tools, IBM has created feedback channels and educational initiatives promoting diversity and inclusion. The company recognizes the need to monitor AI tools throughout the whole life cycle and strives for continual improvement.



8

# Diversity and inclusion

Another essential element of technology ethics is a commitment to diversity, equity and inclusion. IBM's efforts in this space span many parts of the corporation, which together contribute to improving the quality of IBM's products and services, enriching the employee experience and supporting the common good.

The company's 2020 decision to no longer offer general purpose facial recognition software was a prominent example of IBM's commitment to inclusion. Leading up to that decision, IBM had worked with MIT researcher Joy Buolamwini to improve the accuracy of its gender classification software on people of colour, and instituted processes requiring any potential business involving facial recognition software to be discussed with the AI Ethics Board. Announcing the decision to the United States Congress, Arvind Krishna, CEO of IBM, also said: "IBM firmly opposes and will not condone uses of any technology, including facial recognition technology offered by other vendors, for mass surveillance, racial profiling, violations of basic human rights and freedoms, or any purpose which is not consistent with our values and Principles of Trust and Transparency."<sup>26</sup> Another example of inclusive practices in product development is the IBM Digital Health Pass, described at the end of this paper.

Following the murder of George Floyd in the United States in 2020, many technology companies, including IBM, took concrete steps to address racially insensitive and non-inclusive terminology in workplace tools and processes, such as words like "master/slave", "blacklisting" or "whitelisting". IBM began with extensive technical documentation of non-inclusive language in the source code and Application Programming Interfaces. It then created a minimum viable product and scaling plan to replace this language, which included a reporting mechanism for employees when they come across non-inclusive language. A newly established AI operations team continues to examine the company's software.<sup>27</sup>



9

# Education and enablement

Education is a key enabler for the implementation and operationalization of AI ethics principles, and the path to education starts with awareness. While many at IBM were already interested in AI ethics and other ethical issues in technology, others needed to gain more awareness about these subjects. While IBM's diverse, global and multicultural nature is certainly an asset for thinking about the impact of technology around the world, the pluralism of values can also present a challenge. Because of the various types of enablement programmes, which include mandatory annual training, role-based training and community and team engagement, all IBM employees are aware and adhere to the company's values and principles, including those related to AI ethics.

Recognizing this, IBM published "Everyday Ethics for Artificial Intelligence". A guide for improving

developers' awareness of AI ethics, the booklet has five areas of focus: accountability, value alignment, explainability, fairness and user data rights.<sup>28</sup> While some of these areas were refocused or renamed in later editions, others, such as explainability, fairness and user data rights, remain largely the same yet have been expanded in their level of detail.

To make sure all employees know about IBM's commitments, the company implemented AI ethics training through design thinking workshops on bias, trust and other topics. These workshops are currently being delivered to all new and current employees so that everyone is not only aware of these commitments, but also prepared to act on them. In addition, resources such as the "Everyday Ethics for Artificial Intelligence" booklet and the "Learning to Trust Artificial Intelligence Systems" White Paper are also available for free outside the company.



# Partnerships and stakeholder engagements

Stakeholder engagement is an important way for technology companies to understand the potential impacts of their products. IBM regularly engages with research, business, governments and multistakeholder organizations, either institutionally or through individual IBM employees.

The company pursues a broad set of partnerships and stakeholder engagements for several reasons. First, IBM has a commitment and business incentive to help organizations and society in general understand more about AI, educating people about how AI works so that they can better prepare their organizations and society for the future. Second, IBM wants to understand the ways AI is affecting the world so that the company can respond and anticipate where to establish appropriate action. Third, IBM and its employees seek to set expectations, standards and ethical norms for developing, deploying and using AI, so that AI will benefit the world. This will help prevent AI products, technology companies and the technology industry from becoming marred by scandals and other negative events that might ultimately harm both the tech industry and the world. Further, this can reduce the chance of reactive or overzealous regulation hampering the adoption and benefits of AI.

IBM has relationships with several universities, which have led to collaborative technology ethics initiatives. For example, the MIT-IBM Watson AI Lab requires all approved projects to include an emphasis on ethical and societal implications; the [Notre Dame-IBM Tech Ethics Lab](#) focuses on practical models and applied solutions for AI ethics topics; and the Center for Artificial Intelligence, established with the University of São Paulo, Brazil, includes a dedicated initiative on regulatory and public policy issues, such as how people will be affected by AI and automation.<sup>29,30</sup>

IBM has also been involved in government discussion on AI through its nomination of Francesca Rossi to the European Commission's High-Level Expert Group on AI, which defined the AI ethics guidelines that have been used to design the current European regulation proposal.

Beyond IBM's research engagements, the company also provides thought leadership to business leaders on AI ethics. Through the Institute for Business Value (IBV), IBM has conducted business-oriented research on AI since 2014, drawing upon a database of over 20,000 executive-level survey responses on topics such as adoption, value, capabilities and ethics. In 2020, the IBV released a report of research insights,<sup>31</sup> which provided a systematic, quantitative view of general sentiment, relative importance, and organizational responsibility for AI ethics among board members and C-level executives at traditional enterprises – from banks to healthcare providers to retailers and beyond, a somewhat overlooked constituency in the growing literature on the topic.

IBM was a founding member of the multistakeholder organization, the [Partnership on AI to Benefit People and Society](#), which brings together AI corporations with civil society groups for conversations on the best practices for beneficial AI. IBM was also one of the first signatories of the Vatican's [Rome Call for AI Ethics](#) (2020). Further, the company has collaborated on and continues to expand its cooperation in many World Economic Forum initiatives related to AI ethics, including being a founding member of the [Global AI Action Alliance](#), with CEO Krishna as co-chair of the steering committee; Jesus Mantas as a member of the [Global AI Council](#); and Francesca Rossi as co-chair of the [Global Future Council on Artificial Intelligence for Humanity](#). These engagements enable IBM to learn from others in the ecosystem, and others to learn from IBM.

Another form of stakeholder engagement is through the involvement of individual IBM employees in industry organization initiatives, such as the [IEEE's Ethics in Action Initiative](#), the [Future of Life Institute](#), the [AAAI/ACM conference on Artificial Intelligence, Ethics, and Society](#) and [AI for Good](#). IBM is also involved with the [Global Partnership on AI](#), an international multistakeholder organization of governments working to close the gap between the theory and practice of AI.

IBM Research launched the [Science for Social Good](#) initiative in 2016. In this programme, IBM researchers and university PhD or post-doctoral students partner every year with NGOs, social enterprises and government agencies to create solutions for societal challenges using AI technologies (Figure 10). Their research has tackled such issues as treatments for diseases, including

epilepsy and multiple sclerosis; algorithmic fairness for underserved communities; and new methods of drug discovery. These efforts acknowledge both the challenge and extraordinary potential of deploying AI to address key humanitarian and social issues, as driven by the United Nations Sustainable Development Goals.

FIGURE 10 Selected IBM Science for Social Good projects



Source: Science for Social Good projects graphic provided by IBM



Overall, only a multi-dimensional and multi-stakeholder approach can truly address AI bias by defining a values-driven approach, where values such as fairness, transparency, and trust are the center of creation and decision-making around AI.

Francesca Rossi, Global Leader, AI Ethics, IBM, USA<sup>32</sup>

# Summary of progress

Each section of this paper describes the progress IBM has made in advancing technology ethics. This section summarizes the company's major achievements and contextualizes the further work ahead.

It is helpful to view IBM's progress aligned with the company's five commitments in technology ethics:<sup>33</sup>

1. Establishing the internal AI Ethics Board
2. Implementing company-wide education on technology ethics
3. Creating the IBM Trustworthy AI research programme to continuously explore frontier concepts in technology ethics

4. Participating in multistakeholder initiatives on technology ethics
5. Holding regular engagements with important players in the ecosystem to discuss key topics in AI ethics

Although the company is continually looking for ways to evaluate its progress, its advances thus far are visible in the following dimensions:<sup>34</sup>

## 11.1 Influencing policy

As IBM researches and promotes ethical guidelines for technologies such as AI, adoption of these guidelines in public policy and public rhetoric are important measures of success for the company. For example, in the IBM Policy Lab piece "[Precision Regulation for Artificial Intelligence](#)", IBM advocated for targeted policies to regulate AI that focus on specific use cases, industries and affected communities, rather than one-size-fits-all policies. The risk-based AI governance approach allows

for a more precise way of managing use-case examples, such as facial recognition technology. IBM promoted this approach through involvement in multistakeholder dialogues with policy-makers across many channels. EU regulators recently delivered a proposal for the regulation of AI which employs a risk-based policy framework and considers the industry perspective of organizations, including IBM.<sup>35</sup>

## 11.2 Integrating ethics in the AI development pipeline

One view of IBM's progress is how far along the company is at integrating ethics into internal operations. Such integration to develop AI is an important goal for IBM. Fortunately, the company has extensive experience at integrating privacy and security into its processes, programmes and products. Milestones such as the reorganization

that moved the AI Ethics Board and Ethics Project Office under the leadership of the Chief Privacy Officer is an important positive step in this direction. IBM has also mandated that all business units adopt the Ethics by Design methodology. The business units are subject to periodic oversight by the AI Ethics Board.

# What remains to be done

As technology's effects on society continue to grow, IBM works to stay ahead of the ethical challenges that humanity will face by continuing to dedicate research teams to discover and address future challenges. As AI becomes more embedded in the technologies that people use, the boundaries

between people and technology are being redefined. This has created an emerging discipline to responsibly design human computer interfaces, as AI system adoption and human trust of AI systems depend inextricably on each other.<sup>36</sup>



**Behavioural economics prove that humans do what's easy more often than they do what's right. The way user interfaces are designed has a bigger impact on decisions than the actual algorithms behind the systems. When designing ethical and trustworthy AI systems, we must consider both the algorithms and the human-technology user interfaces.**

Jesus Mantas, Senior Managing Partner, IBM, USA

Similar to what they did with AI, IBM and its AI Ethics Board are proactively evaluating the potential ethical impacts of more advanced technologies on society, such as neurotechnology innovation, which has the potential to read and affect human thoughts, emotions and actions. While neurotechnology may have great potential to solve important problems, especially in healthcare, it raises additional ethics concerns.

IBM has a history in innovation, research and thought leadership. As these advances have matured over time, the company is also maturing in its ethical journey. This includes its research efforts, the establishment of trustworthy principles, the formation of a centralized AI ethics governance structure, the development and adoption of tools and education material and initiatives to operationalize AI ethics, within the company and with its partners and clients, as well as its endeavours to promote responsible AI policies throughout the world.

# Conclusion

For many decades, IBM has been involved in responsible technological innovation and digital transformation. The five commitments the company made in 2016 to advance its understanding and effect AI's ethical development have set the stage for how the company is delivering technology solutions for the 21st century. IBM has been working to deliver on these commitments by building the governance, processes, tools and culture to enable trustworthy AI within the company and for its clients. It is a continuation of the company's heritage in ethics values.

The rapid adoption of Fourth Industrial Revolution technologies such as AI will only accelerate as the world continues to navigate through the COVID-19 global pandemic and its aftermath. Continued contributions to the responsible design, development, distribution, deployment and use of technology are essential to create a positive future for the world.

This case study, in collaboration with the Markkula Center for Applied Ethics at Santa Clara University, is shared by the World Economic Forum to present the AI ethics journey, which can encourage and enable organizations to adopt and operationalize ethics in technology in their activities. It aims to promote discussion and evaluation of IBM's methods, tools and experiences to collectively advance the field of responsible use of technology. The World Economic Forum and its partners in this project hope that more organizations not only operationalize ethics in their development, deployment or use of technology, but also share their experience with the global community.

## BOX 2

### Sidebar: The IBM Digital Health Pass

With COVID-19 ravaging the world, and testing and vaccinations uneven and poorly traceable, IBM Watson Health sought an opportunity to help. Using tools developed by IBM Research, the team looked to solve the problem of verifying and tracing health information while at the same time protecting personal data and privacy.

The Watson Health team asked IBM's AI Ethics Board how to operationalize IBM's Principles of Trust and Transparency to transform them into a trusted solution, which became the Digital Health Pass, to make it truly worthy of trust. At the time, Christina Montgomery, IBM Chief Privacy Officer and Co-Chair of the AI Ethics Board, had already described in general terms how these principles ought to apply to technologies relevant to COVID-19, in a blog post where she listed "key guardrails for COVID-19 technologies".<sup>37</sup> They included requiring data transparency, restricting data use to a specific purpose, collecting only the minimum amount of data necessary, building-in privacy and security, requiring humans-in-the-loop for decision-making, and ensuring technologies are lawful, fair, inclusive, and non-discriminatory.

IBM Watson Health collaborated with the AI Ethics Board and designed the Digital Health Pass solution with features that ensure its alignment with ethical values. The Pass is a programme that creates digital certificates to verify vaccination and

testing status, working via an encrypted digital wallet on a user's smartphone. QR codes act as a simple, secure and specific way for information to be transmitted. Privacy is maintained by sharing data very specifically, as determined by users who are given options for what they want to share.<sup>38</sup>

The Digital Health Pass runs on an open blockchain architecture that enhances interoperability while maintaining a decentralized structure that protects its information. The system uses a distributed ledger to validate that only parties with explicit access to a Health Certificate can access it.<sup>39</sup>

IBM highlights several potential uses for the Digital Health Pass, including travel, sports, entertainment, employers and public health. In each case, the Pass helps to ensure that only data necessary for the requirements of the task at hand is shared. For example, if a traveller needs verification of vaccination, then that traveller can share only that specific piece of data with the one organization requesting it.<sup>40</sup>

Attention to ethics has made the Digital Health Pass a better product by aligning privacy and public health to serve the individual, the community and the world. Ethics is vital to this endeavour because even the most effective technical solution will never be used if people mistrust it.

# 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 Fellow, World Economic Forum LLC;  
Seconded from Salesforce

### **Emily Ratté**

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

## Acknowledgements

### **Kathy Baxter**

Principal Architect, Ethical AI Practice, Salesforce, USA

### **Danielle Cass**

Activation Lead, Ethics and Society, Microsoft, USA

### **Kay Firth-Butterfield**

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

### **Ted Kwartler**

Vice-President, Trusted AI, DataRobot, USA

### **Clara Langevin**

Project Lead, Artificial Intelligence and Machine Learning, Centre for the Fourth Industrial Revolution Brazil, Brazil

### **Patrick Lin**

Director, Ethics and Emerging Sciences Group, California Polytechnic State University (Cal Poly), USA

### **Steven Mills**

Partner and Chief AI Ethics Officer, Boston Consulting Group, 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

# Endnotes

1. Hemp, Paul, and Thomas A. Stewart, "Leading Change When Business Is Good", *Harvard Business Review*, December 2004, <https://hbr.org/2004/12/leading-change-when-business-is-good> (accessed 11 August 2021).
2. Krishna, Arvind, "Remarks as prepared: IBM Think Digital Opening Keynote: The New Essential Technologies for Business", IBM Think Digital, 5 May 2020, [www.ibm.com/about/arvind/speeches/05-05-2020](http://www.ibm.com/about/arvind/speeches/05-05-2020) (accessed 11 August 2021).
3. Banavar, Guruduth, "Learning to trust artificial intelligence systems: Accountability, compliance and ethics in the age of smart machines", IBM Corporation, 2016, <https://lexing.eu/wp-content/uploads/2017/06/34348524.pdf> (accessed 11 August 2021).
4. Buolamwini, Joy, and Timnit Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification", Proceedings of the 1st Conference on Fairness, Accountability and Transparency, in *Proceedings of Machine Learning Research*, vol. 81, 2018, pp. 77-91, <https://proceedings.mlr.press/v81/buolamwini18a.html> (accessed 12 August 2021).
5. Raji, Inioluwa Deborah, and Joy Buolamwini, "Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products", AIES, Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, January 2019, <https://dl.acm.org/doi/10.1145/3306618.3314244> (accessed 12 August 2021).
6. Martínez, Cecilia, Ann Gregg Skeet and Pedro M. Sasia, "Managing organizational ethics: How ethics becomes pervasive within organizations", *Business Horizons*, vol. 64, no. 1, January-February 2021, pp. 83-92, <https://doi.org/10.1016/j.bushor.2020.09.008> (accessed 12 August 2021).
7. Skeet, Ann Gregg, "Defining Healthy Organizational Culture", Markkula Center for Applied Ethics, Santa Clara University, 25 October 2019 (updated December 2020), <https://www.scu.edu/ethics/culture-assessment-practice/defining-healthy-organizational-culture> (accessed 12 August 2021).
8. Krishna, Arvind, "IBM CEO's Letter to Congress on Racial Justice Reform", IBM THINKPolicy Blog, 8 June 2020, [www.ibm.com/blogs/policy/facial-recognition-sunset-racial-justice-reforms](http://www.ibm.com/blogs/policy/facial-recognition-sunset-racial-justice-reforms) (accessed 12 August 2021).
9. "IBM's Principles for Trust and Transparency", IBM THINKPolicy Blog, 30 May 2018, <https://www.ibm.com/blogs/policy/trust-principles> (accessed 12 August 2021).
10. IBM, Artificial Intelligence, "Pillars: Explainability", <https://www.ibm.com/artificial-intelligence/ai-ethics-focus-areas> (accessed 12 August 2021).
11. While IBM's other pillars have toolkits to accompany them, the Privacy pillar has a set of resources including policies, principles, products and services which provide information about IBM's work in this area: see IBM website, "IBM Privacy", <https://www.ibm.com/trust/privacy> (accessed 12 August 2021).
12. FICO, "FICO Announces Winners of Inaugural xML Challenge", FICO.com, 31 January 2019, <https://www.fico.com/en/newsroom/fico-announces-winners-inaugural-xml-challenge> (accessed 12 August 2021).
13. IBM Research Trusted AI, "AI Explainability 360", IBM, 2020, <https://aix360.mybluemix.net> (accessed 13 August 2021).
14. Bellamy, Rachel K. E., et al., "AI Fairness 360: An Extensible Toolkit for Detecting, Understanding, and Mitigating Unwanted Algorithmic Bias", ArXiv.org, 3 October 2018, <https://arxiv.org/abs/1810.01943> (accessed 13 August 2021).
15. Sinn, Mathieu, and Beat Buesser, "Adversarial Robustness 360 Toolbox v1.0: A Milestone in AI Security", IBM Research Blog, 13 September 2019, [www.ibm.com/blogs/research/2019/09/adversarial-robustness-360-toolbox-v1-0/](http://www.ibm.com/blogs/research/2019/09/adversarial-robustness-360-toolbox-v1-0/) (accessed 13 August 2021); "Trusted-AI/adversarial-robustness-toolbox", GitHub, 2021, [github.com/Trusted-AI/adversarial-robustness-toolbox](https://github.com/Trusted-AI/adversarial-robustness-toolbox) (accessed 13 August 2021).
16. IBM, "AI FactSheets 360", 2020, [aifs360.mybluemix.net](https://aifs360.mybluemix.net) (accessed 13 August 2021).
17. Arnold, Matthew, et al., "FactSheets: Increasing Trust in AI Services through Supplier's Declarations of Conformity", ArXiv.org, 7 February 2019, <https://arxiv.org/abs/1808.07261> (accessed 13 August 2021).
18. Sattigeri, Prasanna, and Vera Liao, "IBM's Uncertainty Quantification 360 Toolkit boosts trust in AI", IBM Research Blog, 8 June 2021, [www.research.ibm.com/blog/uncertainty-quantification-360](http://www.research.ibm.com/blog/uncertainty-quantification-360) (accessed 13 August 2021); IBM, "Uncertainty Quantification 360", <https://uq360.mybluemix.net/> (accessed 13 August 2021).
19. IBM Cloud Docs, "Building the ground truth", 30 August 2019, [https://cloud.ibm.com/docs/watson-knowledge-studio-data?topic=watson-knowledge-studio-data-build-groundtruth&mhsrc=ibmsearch\\_a&mhq=ground+truth](https://cloud.ibm.com/docs/watson-knowledge-studio-data?topic=watson-knowledge-studio-data-build-groundtruth&mhsrc=ibmsearch_a&mhq=ground+truth) (accessed 13 August 2021).
20. Rudden, Beth, Wouter Oosterbosch and Eva-Marie Muller-Stuler, *Proven concepts for scaling AI: From experimentation to engineering discipline*, IBM Institute for Business Value, September 2020, <https://www.ibm.com/thought-leadership/institute-business-value/report/scaling-ai> (accessed 13 August 2021).
21. Haydock, Michael, Steven Eliuk and Susara van den Heever, *Dealing with the AI data dilemma: The right approach to integration, governance, and tools*, IBM Institute for Business Value, June 2021, <https://www.ibm.com/thought-leadership/institute-business-value/report/ai-data-integration> (accessed 13 August 2021).

22. Plantenberg, Sarah, “Understand Users through scenario mapping”, IBM Garage Methodology, [www.ibm.com/garage/method/practices/think/practice\\_scenario\\_mapping](http://www.ibm.com/garage/method/practices/think/practice_scenario_mapping) (accessed 13 August 2021).
23. Callies, David, and Guy Sharon, “What-if analysis”, IBM Garage Methodology, [www.ibm.com/garage/method/practices/think/what-if-analysis](http://www.ibm.com/garage/method/practices/think/what-if-analysis) (accessed 13 August 2021).
24. World Economic Forum, “Ethics by Design: An Organizational Approach to Responsible Use of Technology”, White Paper, 11 December 2020, [www.weforum.org/whitepapers/ethics-by-design-an-organizational-approach-to-responsible-use-of-technology](http://www.weforum.org/whitepapers/ethics-by-design-an-organizational-approach-to-responsible-use-of-technology) (accessed 13 August 2021).
25. IBM, *IBM 2020 Diversity & Inclusion Report*, pp. 80-81, 2021, [https://www.ibm.com/impact/be-equal/pdf/IBM\\_Diversity\\_Inclusion\\_Report\\_2020.pdf](https://www.ibm.com/impact/be-equal/pdf/IBM_Diversity_Inclusion_Report_2020.pdf) (accessed 16 August 2021).
26. Krishna, Arvind, “IBM CEO’s Letter to Congress on Racial Justice Reform”, IBM THINKPolicy Blog, 8 June 2020, <https://www.ibm.com/blogs/policy/facial-recognition-sunset-racial-justice-reforms> (accessed 13 August 2021).
27. Jones, Dale Davis, and Tim Humphrey, “Words Matter: Driving Thoughtful Change Toward Inclusive Language in Technology”, IBM THINK Blog, 19 August 2020, [www.ibm.com/blogs/think/2020/08/words-matter-driving-thoughtful-change-toward-inclusive-language-in-technology](http://www.ibm.com/blogs/think/2020/08/words-matter-driving-thoughtful-change-toward-inclusive-language-in-technology) (accessed 13 August 2021).
28. Cutler, Adam, Milena Pribić and Lawrence Humphrey, “Everyday Ethics for Artificial Intelligence”, IBM Corp., 2018, <https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf> (accessed 11 August 2021).
29. Arbix, Glauco, and João Paulo Veiga, “AI Humanity – AI in Emerging Countries: Public Policies and the Future of Work”, Center for Artificial Intelligence, 2020, <http://c4ai.inova.usp.br/aihumanity-en> (accessed 11 August 2021).
30. Ministério da Ciência, Tecnologia e Inovações, “Estratégia Brasileira de Inteligência Artificial-EBIA”, [www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/inteligencia-artificial](http://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/inteligencia-artificial) (accessed 13 August 2021).
31. Goehring, Brian C., Francesca Rossi and David Zaharchuk, “Advancing AI beyond compliance: From principles to practice”, IBM Institute for Business Value, April 2020, <https://www.ibm.com/thought-leadership/institute-business-value/report/ai-ethics> (accessed 13 August 2021).
32. Rossi, Francesca, “How IBM Is Working Toward a Fairer AI”, *Harvard Business Review*, 5 November 2020, <https://hbr.org/2020/11/how-ibm-is-working-toward-a-fairer-ai> (accessed 13 August 2021).
33. Banavar, Guruduth, op. cit.
34. More information on how IBM evaluates progress can be found in the company’s *2020 Corporate Responsibility Report*, IBM Corporation, 2021, <https://www.ibm.org/responsibility/2020> (accessed 13 August 2021).
35. “Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts”, European Commission, 21 April 2021, <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206> (accessed 13 August 2021).
36. Mantas, Jesus, “Why artificial intelligence is learning emotional intelligence”, World Economic Forum, 12 September 2018, [www.weforum.org/agenda/2018/09/why-artificial-intelligence-is-learning-emotional-intelligence/](http://www.weforum.org/agenda/2018/09/why-artificial-intelligence-is-learning-emotional-intelligence/) (accessed 13 August 2021); Mantas, Jesus, “Empathic AI could be the next stage in human evolution – if we get it right”, World Economic Forum, 2 July 2019, [www.weforum.org/agenda/2019/07/empathic-ai-could-be-the-next-stage-in-human-evolution-if-we-get-it-right/](http://www.weforum.org/agenda/2019/07/empathic-ai-could-be-the-next-stage-in-human-evolution-if-we-get-it-right/) (accessed 13 August 2021).
37. Montgomery, Christina, “Using Data Responsibly to Tackle a Global Emergency”, IBM THINKPolicy Blog, 18 June 2020, [www.ibm.com/blogs/policy/using-data-responsibly-to-tackle-a-global-emergency/](http://www.ibm.com/blogs/policy/using-data-responsibly-to-tackle-a-global-emergency/) (accessed 13 August 2021).
38. IBM, “Digital Health Pass – Overview”, [www.ibm.com/products/digital-health-pass](http://www.ibm.com/products/digital-health-pass) (accessed 13 August 2021); IBM, “Digital Health Pass – Features” IBM, [www.ibm.com/products/digital-health-pass/features](http://www.ibm.com/products/digital-health-pass/features) (accessed 13 August 2021).
39. IBM, “How IBM Blockchain technology powers IBM Digital Health Pass”, [www.ibm.com/watson/health/resources/digital-health-pass-blockchain-explained](http://www.ibm.com/watson/health/resources/digital-health-pass-blockchain-explained) (accessed 13 August 2021).
40. IBM, “Digital Health Pass – Overview”, op. cit.



---

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](mailto:contact@weforum.org)  
[www.weforum.org](http://www.weforum.org)