WORLD ECONOMIC FORUM

A Holistic Guide to Approaching AI Fairness Education in Organizations

WHITE PAPER

SEPTEMBER 2021

Cover: GettyImages/metamorworks

Inside: GettyImages/metamorworks; GettyImages/Drazen; GettyImages/AndreyPopov; GettyImages/pixelfit; GettyImages/alvarez; GettyImages/anandaBGD; GettyImages/IgorKutyaev

Contents

Disclaimer

The Network of Global Future Councils is an invitation-only community that serves as a brain trust for the World Economic Forum and the world at large. This paper has been written by the World Economic Forum Global Future Council on AI for Humanity. 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 Forum, nor the entirety of its Members, Partners or other stakeholders, nor the individual Global Future Council Members listed as contributors, or their organizations.

Foreword

Kay Firth-Butterfield
Head of AI and Machine Learning,
Member of the Executive Committee
World Economic Forum

Emily Ratté
Project Specialist, AI and
Machine Learning
Manager, Global Future
Council on AI for Humanity
World Economic Forum

Over the past few years, many technology companies – and, frankly, many organizations that had never thought of themselves as inherently technological – have begun to recognize the importance of ethical and responsible development, design, deployment and use of artificial intelligence (AI). This can be attributed to the increased use of AI in many commonplace organizational functions, such as marketing platforms, talent management tools and search engines; and, by extension, the increasingly frequent ethical dilemmas emerging from the way in which these AI applications have been created or used. Many organizations have developed principles regarding AI ethics, such as transparency, explainability, privacy, robustness and fairness.[1]

The Global Future Council (GFC) on AI for Humanity was convened under the mandate of finding solutions to critical issues of AI fairness. Some readers may wonder why this council of experts is primarily focused on fairness when there are many other issues within the burgeoning field of "AI ethics" worthy of being further addressed. With this council convening in 2020 at a time marked by deep economic and social unrest and injustice, the World Economic Forum hoped to shine a spotlight on fairness as an essential part of any future in which AI continues to be developed, deployed and used at scale.

The GFC comprises 24 experts from around the world, who are making advancements in this space. Representing many professional and cultural backgrounds, sectors and industries, the group recognized, and early on came to a consensus on, the multidimensional nature of AI fairness – which precludes any one definition of what "fair" AI looks like. For this reason, the following report outlines a holistic approach to addressing AI fairness education in an organization, which can be adapted to different sector and industry contexts as necessary. The report draws on several collective values including access, equality, equity and transparency.

Not every organization will have the resources necessary to hire a dedicated team of AI ethicists, let alone experts focused on fairness specifically. Developing curricular materials to educate employees on the potential implications of biased or unfair AI, as well as methodologies, tools and practices to address these implications, will require even greater financial commitment and more resources. With this holistic look at the role and impact of different members of a business in addressing AI fairness, we hope to provide options and a North Star for any organization open to improving its practices or creating products with a positive and equitable impact on the larger population of the world.

Executive summary

As organizations automate or augment their decision-making with AI, there is a high risk that the resultant decisions either create or reinforce unfair bias. The negative impact of bias and unfairness in AI does not affect individual victims alone. Organizations that design, develop and deploy AI can face serious repercussions such as brand/reputational damage, negative sentiment among employees, potential lawsuits or regulatory penalties, and loss of trust from all stakeholders, including customers and the general public.

This report aims to address one important part of organizations' approach to AI fairness: educating different teams about the role they play in advancing AI fairness. Holistic learning and education on AI fairness across an organization can drive employees to understand the important role they play in contributing to better, more equitable and more ethical use of AI.

This paper outlines six functions within an organization that have a particular role in operationalizing AI fairness: senior leadership, chief AI ethics officers, managers, build teams (data scientists, developers, product designers, engineering teams etc.), business teams (customer-facing teams such as sales, marketing and consulting) and policy teams. Each section of the report delineates the responsibilities of the team in contributing to AI fairness outcomes, as well as the competencies and training that should be measured and provided to the team to enable them to carry out their responsibilities.

The report also includes recommendations on further efforts needed to improve AI fairness outcomes beyond education, including defining an organization's fairness objectives, creating a supportive corporate culture and hiring diverse teams at all levels and parts of an organization. It ends with a case study on a child-centred approach to AI fairness, to help readers to contextualize the information presented throughout.

1 Introduction

The Fourth Industrial Revolution is blurring the boundaries between the physical, digital and biological worlds. AI is driving this revolution.

Artificial intelligence (AI) has been in existence since the mid-1950s, when John McCarthy, computer and cognitive scientist, coined the term. AI refers to technologies that employ data and algorithms to perform complex tasks that would otherwise to require human decision-making. AI is considered colloquially to be the simulation of human intelligence in machines. However, AI systems today may perform with high accuracy on a given task or dataset, they do not have "general intelligence", or the ability to autonomously comprehend and respond to any decision, particularly in social contexts, in the same way that humans do as they constantly make decisions throughout their day.

1.1 A brief introduction to AI fairness

As a technology that can be used for a number of ends, including increasing efficiency and automating rote tasks, AI is being deployed to some degree in most large-scale organizations. The extent of AI implementation has become a differentiator for many businesses. According to KPMG, 84% of financial services companies state that AI adoption accelerated during the COVID-19 pandemic.[2] However, as AI picks up momentum in business applications across the globe, the question of AI fairness looms large.

AI fairness is a pillar of the larger field of AI ethics, which aims to maximize its positive impact while mitigating the risks of AI to benefit humans and the environment. AI ethics studies the design, development and deployment of AI systems in accordance with agreed-upon values and principles such as data responsibility, privacy, inclusion, transparency, accountability, security, robustness and fairness. While these principles and values may appear in an organization's code of conduct, AI ethics and fairness should be approached holistically, beyond compliance, as a continuous process to improve products and services to better serve both customers and broader society, ensuring AI's life cycle protects human rights and well-being.[3]

Decisions made by AI systems are said to be fair if they are objective with regard to protected indicators such as gender, ethnicity, sexual orientation or disability and do not discriminate among various people or groups of people. For example, an AI-based hiring system may recommend candidates who are more outgoing or extroverted because many extroverted candidates were hired in the past. However, this decision does not take into account whether introverted mannerisms could be a result of cultural differences. This could be an unfair outcome of a technically accurate AI system.

Because an AI system touches many teams within an organization before it is used by a customer or stakeholder – design, data science, developers, marketing, sales etc. – AI fairness cannot be made the responsibility of one team alone. For example, design teams, in their excitement to build a natural language processing model that functions in many linguistic contexts, may overlook issues of access for users with hearing impairments. Similarly, sales teams focused on AI's power to create value might neglect the ethical implications of a sale to an authoritarian government.

This report aims to address one important part of organizations' approach to AI fairness: educating different teams about the role they play in advancing AI fairness. Holistic learning and education regarding AI fairness in an organization's ecosystem can drive employees to understand the important role they play in contributing to better, more equitable and more ethical use of AI.

1.2 Why fairness?

As organizations automate or augment decision-making with AI, there is a high risk that decisions either create or reinforce unfair bias. The problem of bias is not unique to AI. Creating fair and equitable systems is still a work in progress for all societies. This stems from core social issues unrelated to technology, such as structures of power and economy, and the lack of inclusion of heterogeneous perspectives in decision-making. Just as we take steps to address discrimination through education, public policy and regulations,[4] we need to take steps to mitigate against unintended and inappropriate discrimination embedded in AI systems. This is especially important since AI systems, largely designed by homogeneous groups – only 26% of positions in data and AI are held by women,[5] and around 67% of AI professors are white[6] – may amplify the biases of developers, thereby harming exponentially more people.

The negative impact of bias and unfairness in AI does not affect individual victims alone. Organizations that design, develop and deploy AI can face serious repercussions including brand/reputational damage, negative sentiment among employees, potential lawsuits or regulatory

It is against this backdrop that we deliver concerted attention to AI fairness within the larger realm of AI ethics.

penalties, and loss of trust from all stakeholders including customers and the general public.

Just a few years ago, discussions about AI fairness were still mostly conducted in academic and research circles. In recent years, as greater attention has been paid to real-life scenarios and use cases in which individuals have been harmed by algorithmic decision-making, many organizations are working to apply AI fairness research to improve their workforce development and business management processes. This may include using ethical and inclusive design practices in the initial stages of product design, or implementing gatekeeping processes so that AI products are not deployed before they have been checked for fairness.

According to a recent PWC market study, 47% of organizations test for bias in data, models and human use of algorithms.[7] But, according to a 2021 BCG study, executives are broadly overestimating how responsible they're actually being and aren't appropriately measuring their use of AI against practical guidance frameworks.[8] According to BCG's assessment of organizations' responsible AI (RAI), of the companies it surveyed, 14% were lagging, 34% were developing, 31% were advanced and 21% were leading.

BCG also found that organizations are seriously overestimating their RAI progress. When BCG asked executives how they would define their organization's progress on its RAI journey, results indicated: no progress (2% of respondents), had defined RAI principles (11%), had partially implemented RAI (52%) or had fully implemented RAI (35%).[9]

Steven Mills, Managing Partner and Chief AI Ethics Officer at BCG, stated that: *"The results were surprising in that so many organizations are overly optimistic about the maturity of their responsible AI implementation. While many organizations are making progress, it's clear the depth and breadth of most efforts fall behind what is needed to truly ensure responsible AI implementation."*[10]

It is against this backdrop that we deliver concerted attention to AI fairness within the larger realm of AI ethics. We identify four main reasons why AI fairness requires explicit attention from organizations today:

First, harms caused by biased results of AI systems are not trivial, and can affect large groups of people in substantial ways. One type of harm is when a biased AI system allocates or withholds an opportunity or resource from certain groups. If a biased decision was made for consequential decisions for large populations (e.g. access to educational institutions or government grants), the damage will be large. Another type of harm could result if a biased AI system does not work as well for certain groups. A typical example of "quality-of-service harms"[11] is the varying accuracy in face recognition for different ethnic groups, which has a wide-ranging impact. A biased AI system could also reinforce discrimination against certain groups by perpetuating stereotypes.[12] A noteworthy example of this is the use of the COMPAS algorithm, which disproportionately rated black defendants at higher risk of recidivism when compared with their white counterparts.[13]

Indeed, it is notable that the extent of the impact of unethical behaviour by human beings tends to be limited by an individual's relationships or position in a community or organization. The effect of biased or unfair AI systems, however, can be far greater, as software created in one corner of the world can easily be shared, sold and scaled around the world.

AI fairness could go undetected unless attention is paid to it. As humans are inherently biased in many ways, based on factors such as upbringing, education, environment and more, they may unconsciously bake biases into the AI systems they build. Without awareness by the teams building AI models of the importance of AI fairness, if the AI system is deployed, the harms caused may go undetected. Take, for example, the use of AI in resource allocation – unless two individuals from different groups compare what they received, the discriminated individual will not realize that they received less than the other individual. Quality-of-service harms could similarly go undetected for a long time, especially if a discriminated individual subconsciously accepts a lower standard of service because it aligns with other similarly discriminatory experiences. Harms from AI systems that perpetuate common stereotypes are even more insidious as they may go unrecognized in non-diverse companies and communities. Fundamentally, all types of harm caused by AI bias require a deliberate choice by an organization or stakeholder to investigate and mitigate them.

Lastly, and most critically, AI fairness is a socio-technical challenge, and there is no right answer. Fairness is a social construct and has many different definitions in different cultures and for different contexts. Encouraging a diverse and inclusive AI ecosystem is thus all the more crucial to ensure that one definition of fairness does not contradict another, and that the process of defining fairness itself is fair, with under-represented groups at the table leading the conversation. While there are tools that can help to assess fairness through various metrics, AI fairness is not a problem that can be solved simply through technical means. Beyond defining fairness in line with non-discriminatory values, human and child rights, an organization's risk appetite, business objectives and customer expectations, organizations will need to engage with all stakeholders and deliberate carefully on what fairness means for their specific AI use case.[14]

Managing reputation risk is a huge business challenge. Managing AI fairness will reduce risks that companies face down the line which, without proper attention, could severely damage the company's reputation and affect revenue, on top of the harms to stakeholders. Educating company employees on AI fairness provides a critical defence, and has tangible financial value in minimizing the cost of future scandals, regulatory penalties, litigation costs and customer loss.

1.3 Approaching fairness education in a corporation

All personnel involved in the development, deployment and use of AI systems have a role and responsibility to operationalize AI fairness and should be educated accordingly.

Given that AI fairness is a socio-technical challenge, it is not only the responsibility of those in technical roles to address it. All personnel involved in the development, deployment and use of AI systems have a role and responsibility to operationalize AI fairness and should be educated accordingly.

General education on AI fairness
Personnel will not be able to carry out their responsibilities meaningfully unless supported by a corporate culture that shares a common conviction, understanding and awareness about AI fairness. A culture of openness to raise ethical issues, discuss their trade-offs and implications, and act on the outcomes, is essential to ensure that AI fairness issues are not swept under the carpet in order to meet deadlines.

Organizations define what AI fairness means to them, based on their organizational core values and higher societal values.[15] Drawing on this pattern, we propose that the first step in approaching AI fairness is to develop an AI ethics charter with a strong chapter on AI fairness. In doing so, an organization defines what AI fairness means to them, based on their organizational core values. This charter can guide the overarching strategy and decisions in relation to AI fairness, as well as help employees at all levels distinguish between right and wrong decisions or actions.

Importantly, the process by which the charter is written matters. Rather than being created in a top-down fashion by leadership teams, organizations should start this step by enabling conversations, creating structures for enhanced cross-team relationships and communication, and building trust among stakeholders. Employees, contractors, users and customers will be best positioned to raise concerns and offer suggestions to improve a product or system, and deserve to be consulted on – if not to drive – this process. Developing a fair process for defining values is a good place to start for organizations looking to operationalize their commitment to AI fairness.

After due time has been allocated to developing the charter, organizations should dedicate resources to promulgating the principles and guidelines developed in the charter, e.g. by developing a glossary and e-learning module. Resources must be allocated to ensure that the people throughout the organization understand AI fairness and the role they play in encouraging it.

Role-specific education on AI fairness
To turn principles into practice, organizations will also need to assign roles and responsibilities. Education and training will need to be provided so that personnel in different roles are able to carry out their responsibilities and recognize the ways in which they have an impact on AI fairness. To do this, organizations should consider role- and responsibility-specific education on AI fairness. In this paper, we have defined six key functions within an organization that have a particular role in operationalizing AI fairness. These are senior leadership, chief AI ethics officers, managers, build teams (data scientists, developers, product designers, engineering teams etc.), business teams (customer-facing teams such as sales, marketing, consulting) and policy teams. In these teams, multiple competencies are needed to address AI fairness, including but not limited to:

- **Mobilize employees on AI fairness.** Is leadership engaged in and committed to AI fairness? Are teams embracing AI fairness values in spirit and action? Companies must identify where there is support for these initiatives, and which parts of the organization may be less interested in AI fairness. Only with this understanding can companies substantively change their culture to increase awareness of, and attention to, key issues.

- **Identify possible harms to different stakeholders.** Based on the AI use case, teams should consider potential harms that could affect different stakeholders, including direct and indirect users. The potential harm could arise from an error made by the AI system, as well as when the AI system works as intended. In this process, the team identifies the sensitive features and its privileged and unprivileged groups. This kind of education involves guiding employees to empathize and envision how the AI system could cause harm. Teams could be trained to use tools similar to Microsoft's Judgment Call, which draws on value-sensitive design and design fiction to unearth ethical concerns by having employees write fictional product reviews based on different personas, such as a young digital native and an elderly person.[16] Effective tools get teams to discover unexpected potential harms by identifying and relating to different stakeholders.

– **Define fairness objectives that are relevant to the use case.** There are many different definitions of AI fairness, depending on different use cases or cultural contexts. Employees need an in-depth understanding of the various possible AI fairness use cases, such as equal opportunity, demographic parity, equalized odds etc. Teams must learn to apply their AI ethics charter, and evaluate the fairness implications depending on the use case they are facing.[17] Notably, fairness objectives can include technical metrics (e.g. based on probabilistic distributions) as well as social objectives, which might vary depending on local context.

– **Detect and evaluate bias.** Tests to detect bias in an AI system, such as counterfactual fairness assessments, should be required and integrated, based on the fairness objectives defined earlier. All teams should be equipped either with toolkits to detect bias, or the resources to help conduct such tests. It may not be possible to eliminate all bias, and personnel evaluating the AI system will need to assess if the level of bias present is acceptable. Such personnel should be comfortable with technical fairness metrics, as well as qualitative impacts of biased systems. Since there will be trade-offs to consider, personnel involved in evaluation should also have training in risk management.

– **Mitigate bias.** Teams and relevant leadership should be involved in identifying mitigating measures to address this bias. This could include collecting more data to ensure that the data used is representative of the population who make up the end users of the AI model, adjusting data samples of underprivileged groups, or even revisiting the design and purpose of the algorithm itself. Note that fairness intersects with other pillars of AI ethics such as transparency and explainability; while this report focuses on fairness, the other pillars are crucial and deserve further attention.

– **Monitor whether the level of bias changes when deployed.** Teams would need to set up processes to monitor changes in fairness metrics throughout the life cycle of the AI system. This could require investment in proprietary systems that introduce AI governance-by-design, and provide visibility of metrics to enable monitoring of AI fairness metrics – and by extension, training for teams to monitor and investigate irregularities in fairness metrics when the AI system is live. It could also involve educating consulting and account management teams to raise a red flag if they come across non-technical indicators of a fairness issue.

– **Engage stakeholders on potentially biased output.** Personnel involved in engaging stakeholders of the AI system should be trained to be aware of the benefits, risks and limitations of the AI system – particularly for under-represented groups, unexpected users and vulnerable communities – so they know when to alert the product team. They should also be trained to handle customer complaints with great sensitivity and empathy. This process will further help the organization identify if there are any excluded user groups for which the product should be revisited or redesigned.

In addition to these core capabilities, it is crucial that teams themselves are diverse and inclusive. Not only will non-diverse teams face many more obstacles in identifying and mitigating biased outcomes, unrepresentative groups have a far greater likelihood of embedding their own biases into the AI systems they design or develop, or even perpetuating issues of access and digital literacy in the inequitable deployment of AI systems to customers and users. Diverse teams have also been proven to be more adept at pointing out different perspectives and cognitive biases in their team members.

“In addition to these core capabilities, it is crucial that teams themselves are diverse and inclusive.

FIGURE 1 **These are the kinds of interdisciplinary questions that every organization is beginning to grapple with. Answering them requires a holistic approach to AI fairness, and collaboration across different parts of an organization**

Source: Yousif, Nadjia and Mark Minevich, *10 Steps to Educate Your Company on AI Fairness*, World Economic Forum Agenda, 9 June 2021, https://www.weforum.org/agenda/2021/06/10-steps-to-educate-your-company-on-ai-fairness/.

Large enterprises have had institutional compliance boards for decades with the role of monitoring legal, moral and ethical issues of their companies' goals, objectives and actions. The next evolution must be to supplement compliance processes with informed learning and education opportunities relating to AI ethics and fairness. This will initiate a culture change aimed at improving a company's impact on society in general.

Certain roles may cover a range of responsibilities, and there will inevitably be overlaps in the responsibilities of different teams across an organization with regards to AI fairness. This is a good thing, as shared responsibilities will enable an interdisciplinary, multistakeholder approach to addressing AI fairness in an organization. The following sections will break down the importance of AI fairness education in different company roles. The white paper will conclude with recommendations on implementing a holistic approach to AI fairness education in corporations.

2

Senior leadership

Members of the senior leadership team have a crucial role in AI fairness, directing strategy and allocating resources.

Responsibilities
Efforts to prioritize AI fairness will not be effective without the involvement of senior leadership, who can direct strategy and allocate resources towards implementing educational initiatives. Senior leaders play several important roles in ensuring fairness priorities succeed.

Boards of directors are expected to hold management to the highest ethical standards. They establish the governance structure, the principles and the values of the organization and oversee the overall strategy of the business. Board buy-in can help motivate not just one company, but entire industries to devote attention to AI fairness. In addition, issues involving AI fairness can have significant legal and regulatory consequences that the board must be aware of and help prevent.

Given their powerful platform and ability to put pressure on internal and external actors, chief executive officers and executive teams are crucial supporters of any ethics or fairness initiative. While initiatives and directives should not be solely top-down and must reach broad consensus among employees, without executive support, grassroots initiatives may exist but they are unlikely to elicit necessary lasting change in the company or industry culture. This is a symbiotic relationship, as grassroots employee initiatives are vital to ensuring executive directives and lofty mission statements are actually implemented.

Chief executive officers and executive team sign-off can help to legitimize any charter or policy framework, and create broad awareness of an initiative. Executives can reinforce the importance of these efforts by including them in internal- and external-facing presentations and communications, including them in strategic planning processes, and tracking the success and progress of these efforts by holding the management team accountable to metrics that matter for AI fairness. Additionally, the chief executive officer and senior leaders can help ensure that any employee educational initiatives are successful and have the intended outcomes in the longer term of managing risk for the company.

Competencies/training
In order for senior leaders to be highly engaged in creating and enforcing AI fairness tools and procedures, they must be educated about the importance and potential risks of AI bias issues that affect their business. A recent survey by PwC found that 85% of chief executive officers believe that AI will significantly change the way they do business in the next five years.[18] Clearly, there is awareness of the critical importance of AI, but there must also be a deeper understanding of the risks and nuances of AI fairness and the ways in which each organizational function plays a part in ensuring a product does not harm stakeholders.

Leadership teams would also benefit from an educational focus on the "business case for ethics", or the ways in which investing in AI fairness increases sales and brand value in the long term. This may be a counter-intuitive concept for many. Ethics initiatives require an upfront cost for staffing, and may appear to slow down or even halt production of potentially lucrative products, such as facial recognition technologies. However, companies have begun to discover that they can develop a competitive advantage and improve their brand by investing in initiatives related to fairness, trust, transparency – similar to the effect of investing in cybersecurity measures. Particularly in this decade, when hardly a day goes by without a technology company being condemned over for an ethical issue, customers are increasingly likely to choose to partner with or buy from companies whose products they feel are trustworthy, fair and ethical.

Finally, while this is not unique to the technology sector, senior leadership must be well versed on the importance of cultural diversity and the study of culture change. Leaders who understand the ways in which an organization's structure or mission statement affects ethical outcomes will have an advantage when trying to ensure their company aligns with fairness standards.[19]

Educational initiatives in AI fairness for senior leadership can appear to be few and far between, but most universities offer executive education courses on artificial intelligence, and some organizations have begun to tackle this issue. The IEEE's Ethically Aligned Design toolkit also has detailed information on corporate practices.[20] The World Economic Forum has published a full Toolkit for Boards of Directors governing AI systems, and will release a version of the toolkit for C-suite executives in late 2021.

3

Chief AI ethics officers

Many companies now appoint dedicated AI ethics officers, working with staff to ensure AI fairness.

Responsibilities

As technologies advance and the world becomes increasingly interconnected and complex, many organizations are appointing dedicated staff to lead organizational workstreams, such as privacy, data, human resources or AI ethics. Several companies, including Levi Strauss, have established a chief artificial intelligence officer (CAIO) position, and some have even created a C-suite position dedicated solely to AI ethics. This is a relatively new role, tasked with ensuring that the use of AI models is developed with an ethical framework in mind to mitigate the chances of any harms occurring – and, should harms occur, managing the impact on stakeholders. As a member of executive leadership, this employee can also offer sound advice to and build accountability among chief executive officers and boards of directors on the potential unintended aspects of AI and risks posed for the organization. In cooperation with their legal teams, they can help ensure that companies are on the right side of regulatory compliance. Essentially, the role exists to oversee the implementation of many of the governance mechanisms and educational initiatives described throughout this paper. Thus, for companies with the resources, the role can be a powerful tool for implementing AI fairness education.

The name of this role might differ across organizations: "AI ethics lead", "director of AI ethics" or "director of responsible AI" may refer to a similar role.[21] Siemens, for example, has a role that encompasses both ordinary legal compliance and wider ethics issues. Below are a few examples of pioneering AI ethics leaders, who have slightly different titles and mandates:

- **Francesca Rossi**, IBM Fellow, AI Ethics Global Leader, IBM
- **Paula Goldman**, Chief Ethical and Humane Use Officer, Salesforce
- **Natasha Crampton**, Chief Responsible AI Officer, Microsoft
- **Linda Leopold**, Head of Responsible AI and Data, H&M Group responsible
- **Alka Patel**, Head of AI Ethics Policy, US Department of Defense's Joint Artificial Intelligence Center (JAIC)
- **Steve Mills**, Managing Director and Chief AI Ethics Officer, BCG

For the purposes of this paper, we will refer to the general role by the term chief AI ethics officer, or CAIEO.

Due to the multidisciplinary nature of the position, requiring both technical expertise, humanistic (social science) expertise and business know-how, many chief AI ethics officer roles thus far have been filled by ethics or computer science professors. Indeed, CAIEOs must possess the analytical ability to debate complex topics such as algorithmic fairness, bias or exclusion, and integrate diverse company processes for the purposes of operationalizing AI fairness. However, as this is a nascent field, clear credentials have yet to be established. Chief AI ethics officers must also have empathy, as well as communication and relationship-management skills in order to connect across teams, build trust and make people care about new issues. Regulatory or finance experts, activists, philosophers,[22] lawyers or diversity and inclusion specialists may all make skilled chief AI ethics officers.

In fact, Beena Ammanath, Executive Director of Deloitte AI Institute, claims this role must not only be technical, but bring together all of the major areas of concern of data scientists, machine learning engineers and developers.[23] In addition, the CAIEO must hold responsible technology leaders to account in an era of ethics washing, and has a large responsibility to shift the industry towards better and more inclusive practices. At the same time, chief AI ethics officers must understand the business value of investing in ethics and fairness, including costs for product development, implementation and business adoption. Up until now, much of the argument for responsible AI initiatives has centred on risk reduction. The argument is that a robust responsible AI programme can prevent lapses that put companies at risk of litigation, financial losses and reputational damage. While this is certainly the case, and it is important for businesses to consider this, it does not build a strong business case. Instead, CAIEOs should change their mindset and think about responsible AI as a source of value and a strategic differentiator for their companies.

The role may address other key areas including:

- Educating internal employees, stakeholders and company leadership
- Acting as an evangelist and industry spokesperson
- Managing risks and understanding business value/trade-offs of AI fairness; developing strategies for timely identification, prevention monitoring, mitigation and evaluation of AI risks
- Shepherding the development of AI fairness values, an AI ethics charter and other relevant policies; keeping abreast of the regulatory obligations and contractual commitments of the organization
- Playing a vital role in continuous improvement and practice
- Working with the chief executive officer/senior leadership to implement strategy across business units
- Leading by example in day-to-day behaviour and decisions

It is important that this is not an independent effort limited to risk-management practices, but that the CAIEO receives the necessary support and resources of the entire organization.[24]

For example, Microsoft hired Natasha Crampton, who leads Microsoft's Office of Responsible AI, as the company's first chief responsible AI officer. This office puts Microsoft's AI principles into practice by shaping the company's approach to responsible AI. Her team also collaborates with stakeholders within and outside the company to shape new

laws, norms and standards, and to create tools and frameworks that enable product teams to develop AI responsibly, thus ensuring that the promise of AI technology is realized for the benefit of all. Crampton explained that "it is impossible to reduce all the complex sociotechnical considerations into an exhaustive set of pre-defined rules". To overcome this challenge, companies such as Microsoft are developing processes, tools, training and other resources to affirm that their AI solutions reflect their adopted principles.

An alternative to establishing a chief AI ethics officer function is to resource a larger ethics board or council. Enterprises can create a compliance and AI ethics committee with the express purpose of monitoring performance and producing an executive report to senior leadership. This group would also audit corporate compliance based on the corporate code of ethics through monitoring and measurement, as well as develop effective methodologies.

As a larger group, this board or council can include diverse membership from teams across the company. The human resources representative would, then, monitor adherence to the AI ethics charter among their department, and report any potential violations that may occur. The leader of the board or council must then ensure that each member understands and can operationalize AI fairness and ethics across business units.

Competencies/training
The creation of a CAIEO position at a company will enhance competence in AI fairness across the board, and adherence to the organization's prescribed set of ethical standards. The education required for employees in this role will revolve around: 1) understanding how each part of the organization responds to and addresses AI fairness challenges; and 2) understanding how best to engage the entire organization in making the necessary changes to improve AI fairness outcomes.

As an overseer of technical development in AI fairness, the CAIEO must be educated to the extent necessary on key technical features including fairness, explainability, accountability and robustness. They must also have a firm grasp of challenges within data, as trends may shift over time and real-time relevant information must be reflected to address biases.

As the internal leader of AI fairness initiatives across the company, the CAIEO would benefit from educational modules on organizational psychology and how best to structure an organization in a way that will improve AI fairness. In large corporations, the CAIEO will need to combine and align top-down centralized initiatives, such as corporate directives that state how the whole company should detect and mitigate AI bias when building or using an AI solution, with bottom-up initiatives, such as tools specific to a business unit or to an AI solution. Without such coordination and alignment, contradictory messages will be sent both internally and in the company's ecosystem, and opportunities to scale useful AI ethics tools and methodologies to the level of the whole company would be lost.

Finally, as the "face" of the company's efforts in AI ethics, the CAIEO must also learn to convey complex issues of human-machine interaction to company stakeholders.[25] They may also be called upon to speak regarding legal concerns such as criminal activity, personal injury/harms and product liability situations that may arise in the course of AI deployment.

4 Managers

Managers are enablers of AI fairness who build bridges among staff.

Responsibilities
As the bridge between teams and company leadership, managers of the teams described above are important enablers of AI fairness education in companies. In conversation with leadership, managers can advocate for their teams to receive the necessary training. In conversations with their teams, managers can support their employees to make decisions that positively affect fairness, even when they may be difficult or unpopular.

Managers of developer teams, for example, can make sure their teams feel comfortable pausing what they are doing if necessary to consider the ethical implications of a certain feature, even if that pause might mean the team does not meet its original deadline. Managers of sales teams can support their employees to raise a red flag when they think a client may employ the AI product in a way that perpetuates unfairness, even if it means the company does not make a large sale. Furthermore, in companies that already have dedicated strategy and workstreams relating to AI ethics, managers have an important responsibility to translate corporate directives and high-level principles to their employees, ensuring teams understand how the directives apply to their specific team, and how the team can support the company's larger strategic direction.

Competencies/training
AI fairness education *for* managers, then, must include basic information on the importance of AI fairness and the many ways in which seemingly unrelated teams can have an impact on AI fairness outcomes – e.g. a manager of a marketing team may not see their field as related to AI fairness, and should be educated on the potential implications of AI fairness issues as they relate to marketing.

Some parts of manager education should be distinct from what is offered to business teams or build teams, and focus instead on the manager's role as an intermediary between leadership and specific teams, rather than on technical skills. Education for managers could, for one thing, emphasize the need for occasional pauses in product development to ensure the ethical integrity of a product. Education should also detail the business case for ethics, as some managers may be concerned with sales quotas or development deadlines. This education must also be accompanied by formal channels of communication, which managers and their teams can use to report potential ethical issues they encounter.

Finally, corporate strategies regarding AI ethics principles can seem vague and disconnected from the everyday work undertaken by most teams in a company. One way in which managers can increase engagement in and commitment to these principles is to communicate their direct relevance to teams. This might mean the educational programme for managers shares team activities that managers can use to increase engagement and understanding, or that managers use language that makes their employees feel like they have a direct responsibility and role in ensuring AI fairness across the product life cycle. Thus, education for managers should emphasize communication as a vital skill, in addition to general training on the business case for AI fairness and the potential consequences of bias and fairness issues in AI.

5

Build teams

As the people responsible for creating AI models, build teams play a crucial role in AI fairness.

Responsibilities
A "build team" is a team that creates an AI model to be adopted in a service or product. These teams work on all phases of AI design, development and possibly client customization. Build teams tend to include people with different areas of expertise, such as AI designers, developers and data scientists.

Team members tasked to build an AI model are trained on AI approaches (machine learning, planning, search etc.) and on data-related tasks (e.g. data collection, data cleaning, data fusions). However, they may lack knowledge about AI ethics issues and how to assess whether a potential ethical issue is relevant to the AI model or solution being built. Also, these teams are rarely trained on how to deal with such issues should they arise, and how to modify their everyday practices to prevent, mitigate and otherwise account for them.

Fairness in an AI system cannot be achieved by simply testing and fixing up the model after it has been developed, but requires an integrated, multi-step, iterative process that starts from the initial design phase. So, it is imperative that build teams are educated on what the issues are, how to assess their relevance to the products and how to address them.

Competencies/training
The starting point in educating a build team on AI fairness is to make the team members aware of the possible types of discrimination that the AI model may generate, in various deployment scenarios. This information will of course be captured in the organization's AI ethics charter, as described in section 1.3. In this respect, design-thinking sessions are ideal to trigger a culture change wherein teams begin to anticipate the possible negative consequences of what they are building at the earliest stages of design and development.

The second step is to provide comprehensive learning modules on AI fairness: these modules should dive into algorithms, definitions and key terminology, options for metrics and evaluation of fairness in models, and effective methodologies to ensure teams are "speaking the same language". Beyond a basic module mandatory for all, more technical and specific modules should be taken by technical members such as developers, data scientists and engineers.

The third step is to define a methodology by which the team can test and mitigate AI bias. As part of this methodology, teams must ensure they document the work done in addressing these issues. Such methodology should be easy to integrate with current practices, otherwise its adoption will be difficult and unlikely to achieve success. Build teams should also be trained on how to use the software toolkits for testing, mitigating and documenting bias that are required by the methodology.

The methodology used by build teams cannot be created and delivered by leadership alone and simply adopted by various teams. The teams, after their initial education on AI fairness, as well as the customers/users themselves, need to be engaged in refining the methodology and providing feedback. This means that such teams need to be educated in how to describe their work and their AI models to the business executives and strategists who are in charge of defining the methodology. Only through a collaborative process such as this will the methodology be accepted by teams across the company, be relevant and have the right amount of detail to be helpful to technical teams in their everyday work.

6

Business teams

Training business teams on the impact of AI fairness is vital.

Responsibilities
While build teams handle the product through its design and development stages, business teams have the responsibility of getting the AI product into the hands of end users, whether it's clients or the general public. Business teams here refer to employees in positions such as sales, marketing, communications, consulting and more, who handle the product in its deployment and implementation.

These teams rarely undergo technical training in the capabilities of the AI product. Furthermore, employees in business functions responsible for the distribution and implementation of AI products may frequently be working with external companies or individuals in non-technology sectors, who likely have even less understanding of the potential unintended consequences of an AI product. Educating business teams on the potential impact of biased AI, then, is of vital importance, to ensure they are contributing to the ethical use of AI.

Competencies/training
As described throughout this paper, there are opportunities to implement basic education on AI fairness that can help business teams do their jobs. For example, without appropriate education, a marketing team for a human resources solution might over-promise what their product can do because they are unaware of well-known historical instances in which AI products for HR use cases have resulted in the disproportionate hiring of certain racial groups over others. With a minimum of education, this marketing team would be able to manage the expectations of potential clients, which, further down the line, would help companies avoid becoming over-reliant on an algorithm to make fair choices rather than ensuring humans are kept in the loop and in charge of key decisions.

There are also certain competencies that should be trained and evaluated for specific roles. This could include training sales teams on recognizing and managing high-risk use cases that may not be pursued, even if there could be an immense financial opportunity for the company. For example, a large corporation searching for an AI solution to maximize profits in new markets may end up using the AI in ways that ultimately exacerbate greenhouse gas emissions or displace low-income residents of an area. Role-specific training can help ensure the integrity of the AI system is maintained throughout the product life cycle. Consultants or others responsible for working with clients to implement the AI solution may also need training to properly convey the fairness implications to the client teams who will be using the product. If companies simply make a sale, install the software and then leave, employees of the client company may not be aware of potentially proper/improper uses of the product, which could lead to their use having biased and unfair impacts on some users.

7

Policy teams

The policy team assesses AI fairness in the development and use of technology – and the possible impacts, both good and bad.

Responsibilities
Most organizations have dedicated teams with expertise in public policy, government affairs or legal/compliance functions. Referred to in this paper as "policy teams", these teams generally focus on three main tracks relating to AI fairness:

– Technical landscape: understanding the capabilities of AI and ensuring new technologies adhere to existing regulatory standards

– Social landscape: managing the dynamics between enterprises, governments and people; improving the public perception of technologies and managing harms or negative impacts should they arise

– Industry landscape: keeping up-to-date with the ethical approaches of key industry players in technology; working with other companies to address industry-wide concerns about AI

As awareness of AI ethics has increased among industry actors and the public, some governments have taken strides towards thoughtfully and properly regulating data and AI – partnering with multistakeholder groups to co-develop methodologies, or hiring experts in the relatively nascent field of AI governance to equip themselves with the knowledge necessary for future regulation. Of course, approaches to AI fairness and ethics broadly differ based on jurisdictional context. For example, the European Commission proposed a draft regulation in 2021 for AI systems structured on a risk-based framework – distinguishing between levels of risk such as high-risk use cases such as facial recognition requiring dedicated attention from lawmakers, and minimal-risk use cases facing little to no regulation (aside from GDPR).[26] The proposal also bans particularly dangerous uses of AI (use cases of "unacceptable risk") such as social scoring by governments or systems that manipulate human behaviour.

At the same time, because of the slow speed of regulation paired with increasing public awareness of AI ethics issues, many organizations in the private sector and civil society have developed policy and governance frameworks of their own. These "soft" governance approaches have thus far involved creating and aligning business practices to corporate principles or values, implementing ethics checkpoints and greater gatekeeping functions prior to AI deployment, welcoming external AI audits and educating company employees. Of course, companies must be sure to avoid treating fairness as a checkbox; instead, they should make it core to the organization and its values.[27] Policy teams have a responsibility to ensure their company is adhering to regulations, as well as adapting to industry best practices and trends, and minimizing the negative impact that the company has on society. The organization's AI ethics charter can be a useful reference for policy teams in fulfilling this duty.

Competencies/training
Because policy teams may have broad mandates – from laws in specific jurisdictions and national regulations to corporate codes of conduct and trends in industry techniques – it is important for organizations to develop robust policy teams with knowledge of the vast array of ways in which "policy" may be implemented. Policy teams should thus undergo design-thinking courses that help them to consider systemically and holistically the ways in which policy affects the development and use of technology, and the many possible ways in which a technology may harm or benefit society.

In addition, it is essential for policy teams to understand the ways in which a technology may affect society differently based on geographic, linguistic or cultural contexts. AI fairness education for policy teams should include a cultural competency component that will help employees see past the idea that technology is neutral, and understand that AI systems developed from homogeneous or non-representative datasets, and by homogeneous build teams, will not work as well for all groups. Interventions at various points of the AI life cycle – prior to data collection, checks on data collected, throughout development, prior to deployment and after deployment – can help catch issues that, if left alone, may have severe societal and legal consequences.

It is important to note that this involves providing policy teams with enough decision-making authority that they can create processes to enable and incentivize other employees to act ethically. For example, Microsoft has an Ethics and Society team that is tasked with applying a design-thinking approach to technology ethics. However, without coordinating with leaders and decision-making authorities, Ethics and Society would not be able to operationalize their ideas, require employees to undergo certain trainings or implement ethics metrics in employee evaluations.

8 Conclusion

To be effective, AI fairness education must be practised throughout an organization.

While there are innumerable benefits of AI, and thus many reasons why companies are jumping on the bandwagon to design, develop and sell more AI solutions, unethical and unfair AI can have serious repercussions, affecting not only shareholders but all stakeholders in the company ecosystem and community at large. The encroachment of AI into public awareness is relatively recent, and many stepping stones are still required to properly educate corporate workforces on the role they play in ensuring and improving AI fairness. Such education must start with general awareness-raising, but also be role-specific due to the vast diversity of roles and teams that play a part in AI fairness (although they may not realize it).

Organizations and employees that understand the immense potential risks relating to AI fairness, as well as how to engage the entire company in mitigating and addressing those risks, will make more effective AI systems, be better able to realize positive economic and social outcomes, and build trust and brand reputation among stakeholders.

As can be seen in the sections above, AI fairness education should touch many parts of an organization in order to be effective, meaning companies dedicated to fairness must devote substantial resources and time to their commitment. However, it is important for companies to realize that they do not need to have it all figured out from the start. Further efforts in this area could include the following:

- Create an AI ethics charter with a strong chapter on AI fairness
- Create and support an AI ethics board
- Make use of existing tools, which are aggregated in such repositories as the OECD AI Observatory and the AI Fairness Global Library developed by the Global Future Council on AI for Humanity
- Develop AI fairness outcomes learning sessions with customer- and public-facing staff, including basic and role-specific educational modules
- Document and iterate the company's approach to AI fairness and communicate it in staff/supplier trainings and high-profile events, including for customers and investors
- Ensure transparency with regard to the internal use of AI systems to lead by example
- Measure success:
  - Track and report participation and completion of activities, dedicating a standing item in staff KPIs to:
    - Percentage of bias
    - Responsibility and accountability for decisions
    - Quality and source of data
    - Security controls regarding data, models and decision-making
    - Value creation and efficiency
  - Drive multiple metrics to achieve balance between different biases and experiences
  - Measure whether the RAI literacy of the workforce is improving over time
  - Measure whether the organizational culture is shifting to one with ethics at the core
  - Assess whether the reality is aligned with what is stated in corporate reports, regulatory compliance and shareholder and employee meetings and communication

Case study – AI and youth

Many AI ethics principles, guidelines and strategies – both governmental and corporate – advocate for human-centred AI,[28] whereby AI systems and policies should be directed by human rights and aim to serve humanity. This is the correct foundation for the future of AI development. But what extra protections and opportunities are entailed when the human user is a person under the age of 18?

At least one-third of online users are children[29] and they are highly affected by AI-enabled systems, both directly and indirectly. AI systems enable children's toys, video games, chatbots, adaptive learning software and photo overlays on their favourite selfie apps. Algorithmic recommendations steer children's digital lives – from determining what they watch, read or listen to, to how they socialize. Of course, children are also affected indirectly by automated decision-making systems that allocate welfare subsidies, determine healthcare and education access, and assess housing applications. These interactions and impacts are all relevant to how children's rights are upheld or undermined in the digital environment. Yet, in national AI strategies, the rights of children are given very little attention[30] and, in big tech's digital product development cycle, children's rights, needs and use cases are not adequately included. Many Silicon Valley companies choose "strategic ignorance" with regards to their significant user base of adolescents, resulting in digital platforms that are used by but not fairly and responsibly designed for the well-being of the "unseen teen".[31]

For AI systems and policies to be fair for all users, including the substantial child base, UNICEF's draft *Policy Guidance on AI for Children*,[32] makes these recommendations:

– *Ensure capacity building on AI and child rights.* Cutting across the corporate ladder, from senior leadership, chief AI ethics officers and managers, to build and policy teams, all stakeholders should have awareness and sufficient knowledge of children's rights and AI-related opportunities for children's development. Organization-wide awareness of children's rights issues around AI must be supported by a commitment to child-centred AI from the top level of leadership,[33] so that when ethics or development teams raise red flags, they are taken seriously.

– *Support meaningful child participation, both in AI policies and in the design and development processes*. When an AI system is intended for children, or when children can be expected to use the system, or if the system affects children even if they are not direct users, children's meaningful participation in the design and development process is strongly recommended.

Taking a child-centred approach to fair AI is not only the right thing to do in terms of upholding children's rights, it is also good business sense as it capitalizes on customers' demand for trusted and transparent AI solutions for children. Businesses that invest in safe, responsible and ethical AI designed for children can strengthen their existing corporate sustainability initiatives,[34] and mitigate against corporate reputational risks of AI-related harms.[35] As an example of a corporate entity trying to create more child-centred AI, the global retail chain H&M is using UNICEF's Policy Guidance to update its Responsible AI Checklist – a tool that is used by its product teams and product owners, machine learning engineers, data specialists and others involved in developing or using AI capabilities within the company – so that it better includes a children's rights lens.[36]

Contributors

Lead authors

Rohit Adlakha
CXO Revolutionary and Zscaler Executive Advisor, Zscaler

Raja Chatila
Chair, IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, Institute of Intelligent Systems and Robotics, Sorbonne University

Akram Dweikat
Engineering Manager – Network Economics (ML), Deliveroo

Bryan Lim
Assistant Director, National AI Office, Singapore Smart Nation and Digital Government Office

Mark Minevich
Chair, Artificial Intelligence Policy, International Research Centre on Artificial Intelligence under the auspices of UNESCO, Jozef Stefan Institute

Tess Posner
Chief Executive Officer, AI4ALL

Emily Ratté
AI and Machine Learning Specialist; Manager, Global Future Council on AI for Humanity, World Economic Forum

Francesca Rossi
IBM Fellow; AI Ethics Global Leader, IBM

Steve Vosloo
Policy Specialist, Digital Connectivity, United Nations Children's Fund (UNICEF)

Nadjia Yousif
Managing Director and Partner, The Boston Consulting Group

Acknowledgements

Global Future Council on AI for Humanity

Co-Chairs

Raja Chatila
Chair, IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, Institute of Intelligent Systems and Robotics, Sorbonne University

Francesca Rossi
IBM Fellow; AI Ethics Global Leader, IBM

Fellow

Beatrice Dias
Assistant Professor, University of Pittsburgh

Members

Rohit Adlakha
CXO Revolutionary and Zscaler Executive Advisor, Zscaler

Carlos Affonso Souza
Director, Institute for Technology and Society (ITS), Rio de Janeiro

Latifa Al-Abdulkarim
Assistant Professor of Computer Science, King Saud University

Audrey Aumua
Chief Executive Officer, The Fred Hollows Foundation NZ

Ilene Carpenter
Manager, Earth Sciences Segment, HPC AI (Artificial Intelligence) and Mission Critical Systems, Hewlett Packard Enterprise

Akram Dweikat
Engineering Manager – Network Economics (ML), Deliveroo

Arisa Ema
Project Assistant Professor, Institute for Future Initiatives, University of Tokyo

Constanza Gómez Mont
Founder, C Minds; Founder, AI for Climate

Byun Hyung Gyoun
Senior Vice-President, Head of AI/BigData Division, BC Card

Lacina Koné
Director-General, Smart Africa Secretariat

Bryan Lim
Assistant Director, National AI Office, Singapore Smart Nation and Digital Government Office

Alvaro Martin Enriquez
Global Head of Data Strategy, BBVA

Mark Minevich
Chair, Artificial Intelligence Policy, International Research Centre on Artificial Intelligence under the auspices of UNESCO, Jozef Stefan Institute

Safiya Umoja Noble
Assistant Professor, University of California, Los Angeles (UCLA)

Julie Owono
Executive Director, Internet Sans Frontières (Internet Without Borders)

Tess Posner
Chief Executive Officer, AI4ALL

Sara Cole Stratton
Founder of Māori Lab, Aotearoa-New Zealand

Aimee van Wynsberghe
Associate Professor, Delft University of Technology

Steve Vosloo
Policy Specialist, Digital Connectivity, United Nations Children's Fund (UNICEF)

Nadjia Yousif
Managing Director and Partner, The Boston Consulting Group

Jane Zavalishina
President and Co-Founder, Mechanica AI

Endnotes

1. The principles listed here are not exhaustive, but rather describe several that are held in common across most organizations. Other AI ethics principles include sustainability, equality, access and more.
2. Batley, Melanie Malluk, *AI Adoption Accelerated During the Pandemic But Many Say It's Moving Too Fast: KPMG Survey,* KPMG Thriving in an AI World, 2020, https://info.kpmg.us/news-perspectives/technology-innovation/thriving-in-an-ai-world/ai-adoption-accelerated-during-pandemic.html.
3. IBM Institute for Business Value, *Advancing AI Ethics Beyond Compliance from Principles to Practice*, 2020, https://www.ibm.com/downloads/cas/J2LAYLOZ
4. Chatila, Raja and Francesca Rossi, *AI Fairness Is an Economic and Social Imperative. Here's How to Address It*, World Economic Forum Agenda, 22 January 2021, https://www.weforum.org/agenda/2021/01/how-to-address-artificial-intelligence-fairness.
5. World Economic Forum, *Global Gender Gap* Report 2020, 2020, p. 37, http://www3.weforum.org/docs/WEF_GGGR_2020.pdf.
6. Stanford University Human-Centered Artificial Intelligence, *Artificial Intelligence Index Report*, 2021, Chapter 6, p. 12, https://aiindex.stanford.edu/wp-content/uploads/2021/03/2021-AI-Index-Report-_Chapter-6.pdf.
7. PwC, *PwC's Responsible AI Toolkit* [Infographic], https://www.pwc.com/gx/en/issues/data-and-analytics/artificial-intelligence/what-is-responsible-ai.html.
8. Boston Consulting Group, *Many Large Organizations Are Overestimating Their Responsible AI Maturity*, 30 March 2021, https://www.bcg.com/press/30march2021-many-large-organizations-overestimating-their-responsible-ai-maturity.
9. Ibid.
10. Ibid.
11. Fairlearn contributors, *1. Fairness in Machine Learning*, Fairlearn, 2021, https://fairlearn.org/v0.6.2/user_guide/fairness_in_machine_learning.html.
12. Crawford, Kate, *The Trouble with Bias*, speech presented at the Conference on Neural Information Processing Systems (NIPS), 5 December 2017, https://www.youtube.com/watch?v=fMym_BKWQzk.
13. Angwin, Julia, Jeff Larson, Surya Mattu and Lauren Kirchner, *Machine Bias*, ProPublica, 23 May 2016, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.
14. Basis AI, *AI Governance: The Path to Responsible Adoption of Artificial Intelligence,* 2020, https://resources.basis-ai.com/ai-governance-responsible-adoption-of-artificial-intelligence.
15. More on designing an ethical technology organization can be found in this report: World Economic Forum and Markkula Center for Applied Ethics, *Ethics by Design: An Organizational Approach to Responsible Use of Technology*, 2020, https://www.weforum.org/whitepapers/ethics-by-design-an-organizational-approach-to-responsible-use-of-technology.
16. Ethics and Society, *Judgment Call 2019*, 15 February 2019, https://www.youtube.com/watch?v=LXqIAXEMGI0.
17. A beneficial tool is the fairness tree from Aequitas, available here: *Aequitas* [Infographic], Center for Data Science and Public Policy, University of Chicago, http://www.datasciencepublicpolicy.org/projects/aequitas/.
18. PwC, *PwC's Responsible AI Toolkit* [Infographic], https://www.pwc.com/gx/en/issues/data-and-analytics/artificial-intelligence/what-is-responsible-ai.html.
19. See, for example: *Diversity Wins: How Inclusion Matters, McKinsey & Company,* 19 May 2020, https://www.mckinsey.com/featured-insights/diversity-and-inclusion/diversity-wins-how-inclusion-matters.
20. IEEE, *Ethically Aligned Design in Practice, First Edition, Methods to Guide Ethical Research and Design, Section 2—Corporate Practices on A/IS*, 2020, https://ethicsinaction.ieee.org/#series.
21. In a recent article, Tom Davenport discussed the "seven key types of chief data officer (CDO) jobs", noting that one of the variant roles, data ethicist, is "growing in popularity" and "is [focused] on the ethics of data management, specifically on how it's collected, safeguarded and shared and who controls it. There is no doubt that consumers, regulators, and legislators are becoming more concerned about the misuse of data." Davenport, Thomas H. and Randy Bean, *Are You Asking Too Much of Your Chief Data Officer?,* Harvard Business Review, 7 February 2020, https://hbr.org/2020/02/are-you-asking-too-much-of-your-chief-data-officer.

22. To deal with AI-related processes through a human lens, philosophy graduates would be in heavy demand by 2030, employed to "look into AI-related outputs through a human lens". Arakal, Ralph Alex, "Philosophy Graduates Will Be in Great Demand by 2030: Experts", Deccan Chronicle, 5 September 2018, https://www.deccanchronicle.com/nation/current-affairs/050918/philosophy-graduates-will-be-in-great-demand-by-2030-experts.html.

23. Ammanath, Beena, *Does Your Company Need a Chief AI Ethics Officer, an AI Ethicist, AI Ethics Council, or All Three? Positioning Your Organization for Success on AI Ethics*, Deloitte AI Institute, 2021, https://www2.deloitte.com/content/dam/Deloitte/us/Documents/process-and-operations/ai-institute-aiethicist.pdf.

24. There are several interesting new roles/titles that are being developed. Two promising examples are:

    – Ethics sourcing officer: an ethical sourcing officer would lead an ethics team and ensure that the allocation of corporate incomes aligns with the standards set by customers and employees. That person would also investigate, track, negotiate and forge agreements around the automated provisioning of goods and services, to ensure ethical agreement with stakeholders.

    – Chief trust officer: this professional would work alongside finance and PR teams to advise on traditional and cryptocurrency trading practices to maintain integrity and brand reputation.

    These new creations and enterprises' forward thinking are a promising aspect in the development of more in-depth and powerful implementation of ethical AI across the business universe. Pring, Ben, *21 Jobs of the Future: A Guide to Getting and Staying Employed Over the Next 10 Years*, Cognizant Center for the Future of Work, 2019, pp. 12–36, https://www.cognizant.com/whitepapers/21-jobs-of-the-future-a-guide-to-getting-and-staying-employed-over-the-next-10-years-codex3049.pdf?utm_source=perspectives&utm_medium=internal_banner&utm_campaign=perspectives_internal.

25. Gameblin, Olivia, *Brave: What It Means to Be an AI Ethicist*, AI and Ethics, vol. 1, 2021, pp. 87–91, https://link.springer.com/article/10.1007/s43681-020-00020-5.

26. European Commission, *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, 2021, https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206.

27. See more here: Sweidan, Masa, *The Chief AI Ethics Officer: A Champion or a PR Stunt?,* Montreal AI Ethics Institute, 16 May 2021, https://montrealethics.ai/the-chief-ai-ethics-officer-a-champion-or-a-pr-stunt/.

28. Field, Jessica and Adam Nagy, *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI*, Berkman Klein Center for Internet and Society at Harvard University, 2020, https://cyber.harvard.edu/publication/2020/principled-ai.

29. Livingstone, Sonia, Jasmina Byrne and John Carr, *One in Three: Internet Governance and Children's Rights,* UNICEF Office of Research, 2016, https://www.unicef-irc.org/publications/795-one-in-three-internet-governance-and-childrens-rights.html.

30. Penagos, Melanie, Sara Kassir and Steven Vosloo, *Policy Brief on National AI Strategies and Children: Reviewing the Landscape and Identifying Windows of Opportunity*, UNICEF, 2020, https://www.unicef.org/globalinsight/media/1156/file.

31. Lenhart, Amanda and Kellie Owens, *The Unseen Teen: The Challenges of Building Healthy Tech for Young People*, 2021, https://datasociety.net/library/the-unseen-teen.

32. See *UNICEF Policy Guidance on AI for Children*, https://www.unicef.org/globalinsight/reports/policy-guidance-ai-children.

33. Metcalf, Jacob, Emanual Moss and danah boyd, *Owning Ethics: Corporate Logics, Silicon Valley, and the Institutionalization of Ethics*, Social Research: An International Quarterly, 2019, 82(2), pp. 449–476, https://datasociety.net/wp-content/uploads/2019/09/Owning-Ethics-PDF-version-2.pdf.

34. Save the Children, the UN Global Compact and UNICEF, *Children's Rights and Business* Principles, 2012, https://www.unicef.org/csr/css/PRINCIPLES_23_02_12_FINAL_FOR_PRINTER.pdf.

35. Capgemini, *Why Addressing Ethical Questions in AI Will Benefit Organisations*, https://www.capgemini.com/gb-en/research/why-addressing-ethical-questions-in-ai-will-benefit-organisations/.

36. Office of Global Insight and Policy, *Policy Guidance on AI for Children: Pilot Testing and Case Studies: Gathering Real Experiences from the Field*, https://www.unicef.org/globalinsight/stories/policy-guidance-ai-children-pilot-testing-and-case-studies.

WORLD ECONOMIC FORUM

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