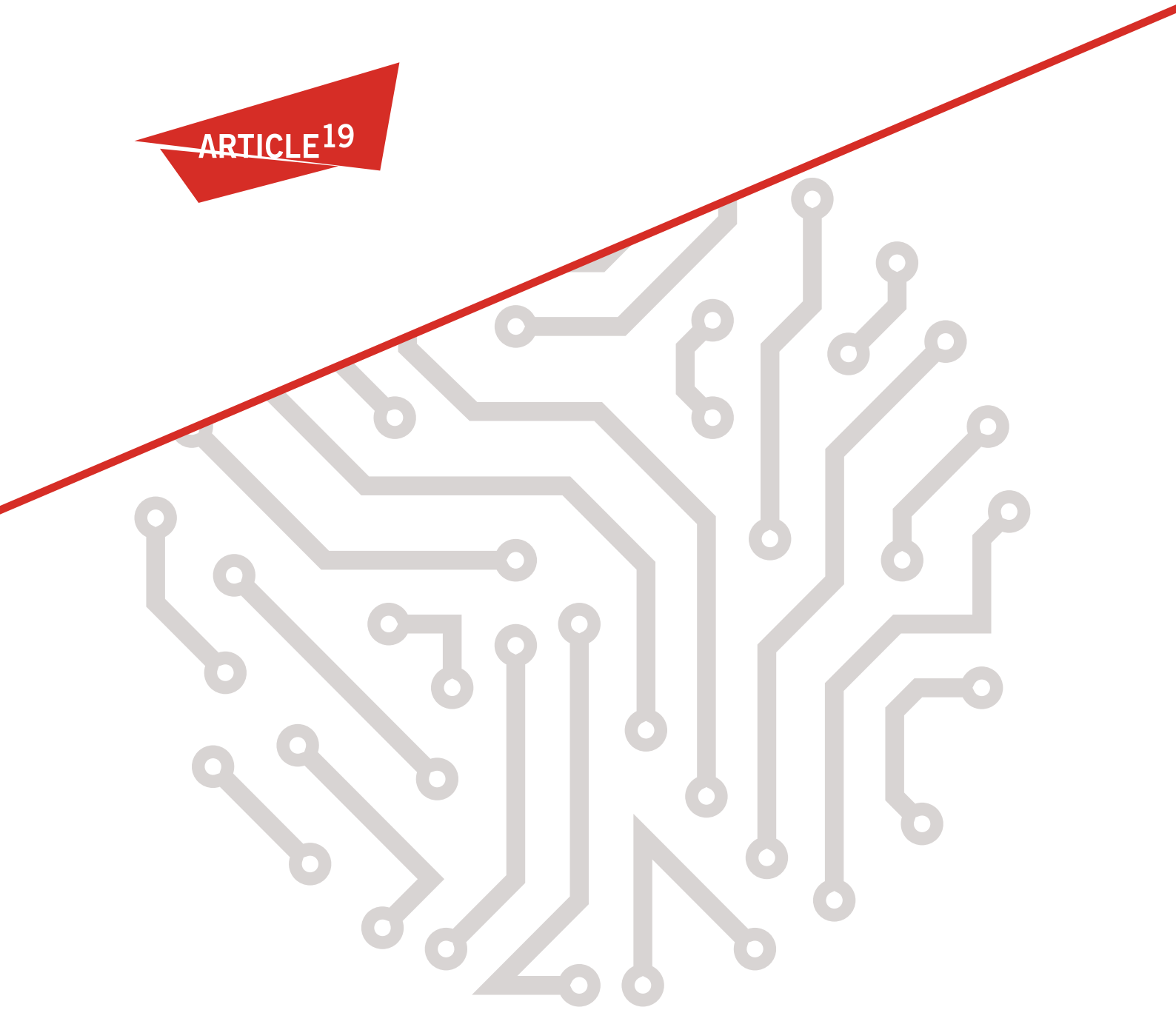
ARTICLE 19



Governance with teeth:

How human rights can strengthen FAT and ethics initiatives on artificial intelligence

April 2019

First published by ARTICLE 19, 2019

ARTICLE 19
Free Word Centre
60 Farringdon Road
London EC1R 3GA
UK
www.article19.org
T: +44 20 7324 2500
E: info@article19.org
Tw: @article19org
Fb: facebook.com/article19org

ISBN: 978-1-910793-42-8

ARTICLE 19 works for a world where all people everywhere can freely express themselves and actively engage in public life without fear of discrimination. We do this by working on two interlocking freedoms, which set the foundation for all our work. The Freedom to Speak concerns everyone's right to express and disseminate opinions, ideas and information through any means, as well as to disagree from, and question power-holders. The Freedom to Know concerns the right to demand and receive information by power-holders for transparency good governance and sustainable development. When either of these freedoms comes under threat, by the failure of power-holders to adequately protect them, ARTICLE 19 speaks with one voice, through courts of law, through global and regional organisations, and through civil society wherever we are present.

Contents

About us

ARTICLE 19 is a global human rights organisation that protects and promotes the right to freedom of expression and information around the world. Established in 1987 in London, ARTICLE 19 monitors threats to freedom of expression in different regions of the world, and develops long-term strategies to address them.

ARTICLE 19 actively promotes human rights-respecting Artificial Intelligence (AI), and investigates the impacts of algorithmic decision making on people's lives. Through policy engagement and research, our published work on AI and freedom of expression thus far includes Privacy and Freedom of Expression in the Age of Artificial Intelligence and a policy brief on algorithms and automated decision making.

We have also provided expert input on related topics and served on multiple committees dedicated to AI and human rights. We have served as invited experts to the UNESCO World Commission on the Ethics of Scientific Knowledge and Technology (COMEST) and several other UN processes, including consultations with various Special Rapporteurs. We have made a submission to the UK House of Lords Select Committee on AI; offered expert input to the Council of Europe committee MSI-AUT; and made several submissions to the AI and ethics initiative of the Institute of Electrical and Electronics Engineering (IEEE), where we also maintain co-chairship of several working groups of the initiative. We hold membership in the Partnership on AI and have given guidance on the development of AI for network management at the Internet Engineering Task Force.

Executive Summary

As artificial intelligence (AI) is increasingly integrated into societies, its potential impact on democracy and society has given rise to important debates about how AI systems should be governed. Some stakeholders have put their focus on building normative ethical principles, while others have gravitated towards a technical discussion of how to build fair, accountable, and transparent AI systems. A third approach has been to apply existing legal human rights frameworks to guide the development of AI that is human rights-respecting through design, development and deployment.

In this paper, ARTICLE 19 considers the ethical and technical approaches in the field so far. We identify the contours and limitations of these parallel discussions, and then propose a human rights-based approach to each of them. The intention behind this paper is to explore how a human rights-based approach can constructively inform and enhance these efforts and present key recommendations to stakeholders:

We call on industry to:

1. Affirm their commitment to the **UN Guiding Principles on business and human rights.**

2. Embed **international human rights standards**, particularly those of protecting freedom of expression and privacy, in the development and deployment of AI systems in a manner that lays out precise obligations and commitments for companies building these technologies.

3. Explicitly invoke a **human rights-based approach** to ethical codes of conduct; identify duty bearers and rights holders.

4. Publicly articulate what safeguards are in place to monitor and evaluate ongoing ethical efforts, and institute **accountability and redressal mechanisms.**

5. **Enhance transparency around ethical AI initiatives**, from the constitution of ethical boards (including their powers and functions), to how ethical principles are internally evaluated, to the impact of these initiatives on businesses.

6. Ensure that AI systems undergo rigorous human rights impact assessments, and create feedback and continuous auditing mechanisms for the same.

7. As a continuation of the above, **meaningfully engage with civil society and academia** at each stage of the process, to cultivate constructive criticism as part of internal deliberation processes.

We call on states to:

1. Reaffirm their **commitment to international human rights standards**, particularly those protecting freedom of expression and privacy, and let this commitment guide their specific national AI strategies.

2. Ensure that AI systems in the public sector undergo adequate **human rights impact assessments**, due diligence, and continuous auditing. These are not systems that can simply be rolled out. They should instead be tailored to the exact context and use for which they are intended.

3. Root the design, development, and deployment of AI systems in **constitutional guarantees and human rights standards**.

4. Hold AI systems to **accountability, responsibility, and constitutional standards without dilution or exception**.

5. Ensure that national and international efforts around AI are equally informed by **human rights concerns, constitutional standards, and the public interest** as they are by industry concerns.

6. Establish necessary infrastructure and resources to provide remedy for human rights violations caused by AI systems.

We call on civil society to:

1. Work towards greater **transparency and accountability** of mechanisms (such impact assessments, the Universal Periodic Review, and rights-based mechanisms at national levels) and institutions (like the Human Rights Council, General Assembly, Special Rapporteurs, national and regional institutions that focus on compliance with and enforcement of human rights) that make up a rights-based approach.

2. Advocate for the promotion and protection of human rights in the context of AI systems in a manner that is also informed by **technical considerations and limitations.**

3. Actively engage with ethical initiatives to integrate strong human rights language and protections into their substantive and procedural functioning.

I. Introduction

As artificial intelligence (AI)[1] has demonstrated its power to revolutionise fundamental systems of communication, commerce, labor, and public services, it has captured the attention of the technology industry, public officials, and civil society. The potential of AI to perform tasks with speed and at scale beyond human capability has fueled great excitement. AI systems are already deeply embedded in our everyday lives - from helping us navigate through morning traffic to offering up the day's news, to more nefarious uses of systems for surveillance,[2] warfare,[3] and oppressing democratic dissent.[4] Yet many of the most powerful stakeholders in the field have only just begun to consider the impact of AI systems on society, democracy, rights, and justice.

Technologists, researchers, companies and governments are grappling with the challenges that AI poses for society, democracy, and governance, while simultaneously trying to develop frameworks to guide this technological development. Thus far, the solutions that different stakeholders have put forth to tackle problematic uses of AI and reckon with their unintended consequences are insufficient, as we will show in the pages that follow.

In these processes, at least three distinct approaches to AI standards and principles have emerged. Some stakeholders have put their focus on building normative ethical principles, while others have gravitated towards determining what fair, accountable, transparent AI systems look like from a technical standpoint. A third approach is working to apply existing legal human rights frameworks to guide the development of AI. Our aim is to explore how these conversations can constructively inform one another.

In this paper, we analyse two approaches to understanding and addressing the societal impacts of AI: 1) a normative approach, with a focus on ethics; and 2) a technical approach, with a focus on fairness, accountability, and transparency (FAT). We identify the contours and limitations of these approaches, and then demonstrate how a human rights-based approach can strengthen them. Finally our analysis concludes with concrete recommendations to stakeholder groups.

2. The societal impact of AI: two approaches

2.1 The normative approach: Ethics initiatives

AI systems raise myriad questions for society and democracy, only some of which are covered[5] or addressed by existing laws.[6] In order to fill these perceived gaps, a vocal group of governments, industry players, academics, and civil society actors have been promoting principles or frameworks for ethical AI.[7] While there is an abstract awareness of what "ethics" generally means, there is no precise or shared understanding of the term. It has been subject to multiple interpretations by various stakeholders, and is often defined by industry actors who use it on a case-by-case basis.

Proponents of this normative ethical approach believe that it enables stakeholders to identify opportunities that are socially acceptable or preferable, while at the same time potentially averting costly mistakes by elucidating what is socially unacceptable. As Floridi et al write, *"With an analogy, it is the difference between playing according to the rules, and playing well, so that one may win the game."*[8] Ethics initiatives allow conversations to transcend a single jurisdiction, and encourage input from various stakeholders to crystallise high level commitments to what AI should look like.

At the time of writing this paper, at least 25 countries have published national AI strategies[9] and ethical task forces have cropped up around the world. The European Commission's High Level Expert Group on AI has laid down "Ethics Guidelines for Trustworthy AI" focusing on respect for human autonomy, prevention of harm, fairness, and explicability.[10] Companies are constituting ethical boards, publishing ethical principles for AI, and taking part in multi-stakeholder initiatives in this space as well. Technical organisations like the Association for Computing Machinery (ACM) and the Institute for Electrical and Electronics Engineers (IEEE) have published ethical principles for autonomous systems. And academic[11] and civil society actors have engaged via government consultations[12] and multi-stakeholder forums like the Partnership on AI (PAI).[13]

This plethora of initiatives underlines the need for a framework to discuss the desired impact of new technologies on society - it pushes for an articulation of what *ought* to be done. The current debate around AI and ethics is rich, multi-disciplinary, and takes various forms, including ethics boards, principles, and public statements. ARTICLE 19 recognises the importance of these efforts, but also believes that there is more reason to be critical than accepting of ethics initiatives, as we will discuss in the next section.

2.1.3 Critical gaps in ethics initiatives

A primary reason to be critical of ethics initiatives in isolation is that they, more often than not, are **not actionable**. They do not afford mechanisms that lead to tangible change. The various principles developed by industry and states have, as of yet, **failed to develop strong accompanying accountability mechanisms**. They lack concrete and narrowly defined language,[14] independent oversight or enforcement mechanisms, and clear transparency and reporting requirements. This means that no matter how laudable the principles are, there is no way to hold governments or companies to said principles. The general lack of transparency mechanisms leaves no pathway for other stakeholders to know whether or not companies and governments are complying with their own principles. And in cases where non-compliance is revealed, there are inadequate mechanisms to hold companies and governments accountable for their wrongdoing.

For instance, after Google received pushback from its own employees surrounding Project Maven, a partnership with the US Department of Defence to improve drone targeting using AI, the company published a set of AI principles that elucidated its commitment to ethics, and made a public pledge to *refrain* from building certain types of technology.[15] But Google has not disclosed to what extent these principles are embedded in concrete work in the company, and there has been no demonstrable change in how the company has altered its internal decision making processes. This is particularly worrying because in the case of public-private partnerships such as Project Maven, the accountability that governments otherwise owe the public is potentially diluted by the use of technology built behind closed doors and vague, non-binding commitments.

This lack of accountability can even take a toll when AI systems are still being developed. Recently, McNamara et al conducted a study where software engineers were explicitly instructed to consider the ACM code of ethics as they were developing new products. The study found that the code of ethics had no "observed effect" on their work.[16] This suggests that ethical codes of conduct cannot be a solution in and of themselves, unless accompanied by mechanisms for compliance and accountability.

Lack of transparency in these initiatives is also of foremost concern. While transparency is a principle that is recommended in multiple ethical frameworks, it ironically is a feature that most ethics initiatives lack. There is a culture of black box decision-making in the conversation around AI and ethics, in how ethical boards are constituted and how principles are developed. When Microsoft constituted the AI and Ethics in Engineering and Research Committee (AETHER), to make recommendations on what AI technologies the company should deploy, the company claimed that "sales" were cut off on the committee's recommendation.[17] Yet the extent, subject, and details of intended sales and nature of deliberation remain a mystery.

This approach also does not account for **enforcement or redressal mechanisms**, nor does it contemplate **the duty of companies to act in certain ways**. Furthermore, it fails to provide mechanisms by which consumers, public interest advocates, civil society, and other affected individuals can have agency in the event that companies fail to meet their own ethical standards.

For instance, following scrutiny over its algorithmic systems,[18] Facebook took several steps towards engaging with ethics initiatives. The company backed an ethical AI institute at the University of Munich[19] and became a founding member of the Partnership on AI, a multi-stakeholder group that aims to "study and formulate best practices on AI technologies, to advance the public's understanding of AI, and to serve as an open platform for discussion and engagement about AI and its influences on people and society."[20] Facebook even signed resolutions calling for the development of ethical principles in the US Congress.[21] Yet at the same time, recent research shows that Facebook discriminates on advertisement delivery on the basis of gender and race,[22] and has also been charged with housing-related discrimination.[23]

Facebook's efforts towards ethical AI had no demonstrable bearing on its practices. Indeed, these partnerships and other loose commitments did not create any requirement for the company to follow through on these ideals, nor did they provide mechanisms to hold the company to account.

As the outcomes of Project Maven show, these initiatives become even more complex when they are deployed in public-private partnerships, where government agencies procure privately developed AI systems. In these agreements, sole reliance on ethical frameworks (as opposed to legally-binding constitutional or human rights frameworks) **dilutes state accountability and rights-based obligations**.

As discussed above, there are various cases of governments working together with industry to improve surveillance of dissidents, precision targeting in drones, or facial recognition software for law enforcement purposes. These partnerships regularly take shape in the absence of safeguards or meaningful oversight.

For example, Google's partnership with the Chinese government to develop a censored search engine (known as "Project Dragonfly") would have excluded search results that were viewed as politically "sensitive" by the Chinese government.[24] This represents an egregious violation of international human rights standards on freedom of expression and information, and also violates Google's own ethical principles on AI.

In both public-private initiatives, and corporate-driven efforts, the debate on AI and ethics is **disproportionately influenced by industry initiatives and corporate aims**.[25] Even though a variety of actors are developing ethical frameworks, concerns

from civil society and academia struggle to get industry support, and even in multi-stakeholder settings, are easily diluted.

When civil society is invited to partake in deliberation around ethical AI, the division of seats at the table is not equitable. Ethical initiatives within industry are more often than not opaque to civil society, with most ethical boards and codes of conduct being developed and deliberated exclusively in-house. What is more surprising is that this trend continues even in ethical initiatives convened by governments. For instance, during deliberations at the European High Level Expert Group on Artificial Intelligence (EU-HLEG),[26] industry was heavily represented, but academics and civil society did not enjoy the same luxury. And while some non-negotiable ethical principles were originally articulated in the document, these were omitted from the final document due to industry pressure.[27]

National governments have followed a similar pattern. In India, an AI task force to create a policy and legal framework for the deployment of AI technologies was constituted without any civil society participation.[28] In the United Kingdom, the Prime Minister's office for AI has three expert advisors - one professor of computer science and two industry representatives.[29]

When we look at efforts led by companies, certain stakeholders appear to use ethics initiatives as an **alternative or preamble to regulation**. This approach can carry dangerous consequences for human rights and the public interest. In proposing ethical frameworks or principles to avoid regulation under the guise of encouraging innovation, stakeholders seek to achieve precisely what is discussed above, a practice Ben Wagner has termed "ethics washing."[30] They affect a veneer of "being ethical," yet they have no mechanisms of accountability with which to comply,[31] and thus they face no consequence for their actions. Some may advocate for ethics as a preamble to regulation, arguing that it is too soon to prescribe regulation addressing AI.[32] But in multiple cases, this has proven to be a strategy of simply buying time to profit from and experiment on societies and people, in dangerous and irreversible ways.

Ethics initiatives tend to frame their discussion of AI in terms of companies doing "the right thing," in accordance with high level ethical principles or standards. But we have seen ethical codes of conduct become a smokescreen for doing "the right thing," even when there is no clear understanding of what "the right thing" is, or how to measure it.

For example, Google's ethical principles laid out its aspirations to build AI systems that are socially beneficial, and also to avoid creating or reinforcing unfair bias. A few months later, the company constituted an ethics board, including individuals[33] who demonstrably contradicted the basic assumptions behind Google's ethical principles. This made the principles and actual company practice fundamentally

incongruous, and Google dissolved the board just days later, following public and internal pushback.[34] Google's AI principles, therefore, have no teeth - they do not preclude the company from violating its own principles because they create no obligation or duty in the first place. The link between ethical aspirations and industry duty is weak at best, and non-existent at worst.

Another problem with trusting companies to do "the right thing" comes from their lack of understanding of the **societal impacts of technology** and appropriate ways to deal with them. For instance, in April 2018, in his testimony before the United States Congress, Facebook CEO Mark Zuckerberg revealed the company's increasing reliance on AI tools to solve problems of hate speech, terrorist propaganda, election manipulation and misinformation. But research and media reports have shown that AI tools are ill-suited to do this work - they are not technically equipped to understand societal nuances or context in speech, and often make the problem worse.

In fact, prior to the 2017 escalation of military attacks on Rohingya people in Myanmar, local activists gathered substantial evidence that Facebook was automatically censoring the word "kalar," a derogatory local slur used to refer to Rohingya Muslims. By simply flagging the word as problematic, without accounting for the immediate context in which it was being used, this step led to the censorship of numerous posts in which people attempted to discuss use of the term, its history, and efforts to curb hate speech in the country. Meanwhile, users who wanted to use the term as an insult simply opted for an alternative spelling. All told, Facebook's effort to deploy AI in order to reduce hate speech in this volatile political environment resulted in censorship of legitimate speech and had no demonstrable effect towards curbing hate speech.

Finally, on a more granular level, it is also important to note that while ethical principles are put forth as aspirational goals, even when well-intended and narrowly tailored, there has seldom been guidance on how to provide **balancing mechanisms** for conflicting principles.

2.2 The Technical Approach: Fairness, Accountability and Transparency (FAT)

Widespread use of AI systems on society has brought to the fore concerns around discrimination,[35] injustice,[36] the exercise of rights[37], amongst others. A community of researchers, academics, and scientists have been working to address these issues by developing AI systems that are fair, accountable, and transparent (FAT). This particular field of technical work has grown over decades, and pre-dates the current resurgence in AI-focused ethics initiatives.

The notion of fairness in machine learning (the most popular subset of AI techniques) is arguably the most prominent topic in the field today, with researchers and practitioners attempting to articulate what fairness entails and in turn, operationalise these learnings at the time of deployment.

Fairness is a normative concept and can be defined in multiple ways.[38] Fairness pushes technical thinking beyond just optimisation, to consider the actual impact of AI systems and their implications for society. But this is not without complication. Sometimes one type of fairness can be more appropriate than the other, necessitating a tradeoff. This balancing act requires some guidance. There is also a question of tradeoffs between fairness and other values - a perfectly fair system may not be very accurate, and accuracy can sometimes undermine fairness.[39]

The ability of AI systems to be at once invisible, opaque, and inscrutable also gave rise to efforts to make them accountable and transparent. This is especially important given the increasing use of AI systems in the public sector for consequential decision-making. The overarching assumption here is that transparency affords enough insight into a system, and in turn enables explanation, redressal, appeal, and accountability to follow. The precise feasibility of this approach is the subject of ongoing debate.[40] Some experts believe that peering inside black boxes could lend some insight into the inner functioning of a system, which could then lead to the ability to hold them to account.[41] Others believe that given the evolving nature of AI systems, transparency on its own will not help.[42]

Calls for accountability, however, go beyond this assumption. AI systems are increasingly used to carry out consequential decision making and deliver services that were once the responsibility of states. Accountability is essential in relationships with power differentials, making this question both technically and legally crucial.

The conversation around FAT has been led by academia for many years, and this is still true today. The Association for Computing Machinery Fairness, Accountability, and Transparency (ACM FAT*) in ML[43] is perhaps the most popular venue for FAT work in the field. Some recent papers from the event included work on improving fairness of facial recognition algorithms,[44] distinguishing between fairness and bias,[45] and a study on bias in news and fact checking. [46]

Industry has also engaged with the idea of fairness. In 2018, Accenture published a fairness toolkit to help businesses work towards fair outcomes in the process of deploying AI systems.[47] Spotify and Microsoft researchers recently presented work on the challenges they face when trying to implement fairness on a daily basis for technical experts working on FAT issues.[48] FAT work is also carried out in industry consortiums like the PAI and IEEE, which often pave the way for wider stakeholder engagement with these issues. Venues such as the PAI offer dedicated working

groups on FAT issues, and have strong civil society participation. Other venues like ACM FAT* are still very much technical, leaning strongly towards academia.

2.2.3 Critical gaps in the FAT approach

The design, development, deployment, and assessment of AI systems can be profoundly affected by work towards fairness, accountability, and transparency. While this was traditionally a purely technical field of research, there is growing acknowledgement of the need to bring in expertise from disciplines beyond computer science and mathematics. Experts from law, social science, philosophy, and other disciplines are beginning to engage with the field, and to inform and enhance these concepts.

Even as this field pushes the limits of our current understanding of fair, accountable and transparent AI systems, ARTICLE 19 believes that some concerns with the current approach remain.

First, technical work needs to **imbibe socio-political awareness** in a way that is both meaningful and competent. Many AI systems currently in use make assumptions about social, cultural, and political values that reflect a lack of worldly expertise. Optimising for a particular type of output is not enough if the surrounding societal and historical context is not brought to bear in tandem with technical considerations. No matter how precise a particular technical definition may be, an accurate technical output does not necessarily generate a fair socio-technical outcome for the user or people affected by the technology.[49]

Consider how a credit scoring model might be trained to operate. How could such a model affect historically disadvantaged populations? It is possible that, when studied in isolation, historical discrimination of the past could easily translate to future discrimination. Recognition of structural and systemic inequalities, constitutional guarantees of affirmative action (where applicable), and the responsibility to correct past practices would mean building a technology that transcends this view. This has been recognised by the community itself and is slowly being addressed. Some of the most recent literature in this field attempts to learn and further develop these concepts by borrowing from political philosophy,[50] social sciences,[51] and even the fields of education and hiring.[52]

But significant questions remain: Which values are embedded? How are technical experts positioned to understand them? What shared terms exist around these values, and what are possible ways to codify them?

Second, the FAT approach does not structurally engage with **responsibility, rights, duties, or harms of technical systems** in an actionable way. This it shares in common with the ethics initiatives discussed in the previous section. The FAT

approach does not necessarily articulate responsibility, harm, or expectation of fair treatment to ensure that these goals are met.

For example, in building a predictive policing algorithm, beyond articulating ways in which it is accountable, and the constraints taken into consideration to ensure it will not be unfair or discriminatory, the development of these systems should engage with certain key questions: Who can demand accountability of these systems? Who are the people to whom fairness and transparency is owed? Which authorities and institutions will be accountable?

Also similar to the ethics approach, the FAT approach does not address the tangible effect of technical systems on the exercise of rights, or on people's lives. For instance, there has been recent pressure on technology companies like IBM and Amazon (among others) to ensure that facial recognition systems have equal accuracy rates between vulnerable and dominant racial groups.[53] Ensuring this could satisfy some definitions of fairness, yet it does not engage with the broader question in play. Are facial recognition systems a threat to the exercise of fundamental rights such as privacy and free expression? Does a perfectly fair facial recognition system -- one that is equally and similarly accurate across demographic groups -- have a disproportionate impact on vulnerable groups that have historically been subject to surveillance? Should the faces of individuals from these groups be used in training datasets? What implications does this have on their autonomy and privacy?

Relatedly, the FAT approach has been built to ensure that systems are fair, transparent and accountable, but it does not empower individuals or surrounding institutional mechanisms to challenge the decisions that these systems make. By sidestepping the question of rights and duties, affected individuals and oversight authorities cannot meaningfully challenge systems that fail them. Take for instance New York's Automated Decision Systems Task Force that examines the use of automated decision-making systems by the city to prevent bias or other harms. The task force has adopted principles of equity, fairness and accountability,[54] and has leading AI scientists and researchers among its members. Yet a full year after it was constituted, the task force is unable to carry out its duties because there are no accompanying institutional mechanisms, oversight powers, or rights to investigate uses of automated decision-making systems.[55]

3. Towards a human rights-based approach

Alongside technical and ethical approaches to addressing the societal impact of AI systems, there has been a third approach, focused on law and regulation. While a handful of states have begun to explore legal frameworks for governing AI, various stakeholders have turned to international human rights instruments as a way forward.

There have been preliminary attempts to begin regulating AI at national and regional levels. In the EU, for example, the General Data Protection Regulation (GDPR)[56] articulates a few rights with respect to automated decision making. Under the GDPR, data controllers (typically companies or state entities) are required to provide data subjects with information about "*the existence of automated decision making, including profiling...and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject*."[57] The GDPR also provides that individuals have the "*right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her*."[58] and guarantees that data subjects can seek human intervention and contest the decision.[59] The extent and impact of these provisions are the subject of ongoing debate, and are sometimes referred to as the "right to explanation."[60] In the United States, an Algorithmic Accountability Bill introduced in the Senate in April 2019 contemplates impact assessments for automated decision systems and data protection.[61]

Given current isolated regional and national attempts at regulation of AI systems, the relevance and importance of internationally recognised, legally binding human rights at the national and international levels cannot be overstated. There has been substantial work on the intersection of AI and human rights so far, particularly around freedom of expression, privacy,[62] non-discrimination, and governance.[63] The Toronto Declaration on protecting the right to equality and non-discrimination in machine learning systems,[64] launched in May 2018, aims "*to underline the centrality of the universal, binding and actionable body of human rights law and standards, which protect rights and provide a well-developed framework for remedies*."

More recently, over 100 organisations signed a statement focused on civil liberties concerns regarding the use of pretrial risk assessment tools.[65] The United Nations also has weighed in on the human rights debate: UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression presented a detailed report on the human rights impact of AI systems to the General Assembly.[66]

Some national governments also have adopted a human rights-based approach. In July 2018, the Australian human rights commission launched a three-year project to understand the human rights impact of AI.[67] The governments of Canada and France are steering an international study group aimed at human centric artificial intelligence, using human rights as one of the anchors of the study.[68] There has also been work highlighting the human rights obligations of businesses in the context of AI.[69]

Having discussed the current deliberation around societal impacts of AI, we now wish to propose a way forward. ARTICLE 19 believes that current deficiencies discussed above -- the lack of enforcement, accountability, meaningful redressal, and individual empowerment -- could be constructively supplanted through a human rights-based approach.

The human rights-based approach identifies rights holders (people who use or are affected by technologies) and duty bearers (companies or governments deploying said technologies). It is a universal set of principles, has binding effect, and is based on the rule of law. It draws on an internationally recognised system of law that defines both business and state responsibility and the specific standards they must adhere to. This means that rights, reasonable restrictions, their status under law and implementation in practice, are anchored in a system that is verifiable, specific, and detailed. This international system has grounding in law, is based on commonly understood language and affords procedures and institutions that can help ensure that duty bearers meet their obligations, and that rights holders have recourse to effective remedies.[70]

What makes a human rights-based approach particularly important in the context of AI is the pre-existing focus on the use of privately developed technology by states. The nexus between states and businesses is explicitly addressed in the United Nations Guiding Principles on Business and Human Rights (UNGPs).[71] These safeguards recognise the possibility of states shirking their obligations under international law by contracting with private actors, and they provide mechanisms to ensure the protection and fulfillment of human rights obligations. It is crucial to note that **human rights are both a legal and ethical standard**.

In this section, we will contemplate how a human rights-based approach can benefit existing technical and ethical conversations discussed above, and in turn also identify ways in which the converse is also true, i.e. what aspects of the FAT and ethics conversations can inform existing human rights approaches to AI.

3.1 A human rights-based approach to Ethics

In our analysis, our primary concern with ethics initiatives was their common lack of accountability and enforcement mechanisms. Some appeared to be little more than efforts to skirt regulation, while others, though perhaps well-intended, did not have proverbial teeth: They put forth admirable goals, but offered little if any mechanism of accountability or enforcement. If a human rights-based approach were brought to bear here, it could complement ethical principles with an enforcement mechanism, drawing from the UNGPs, which offer comprehensive guidance on how to make industry practices more concretely accountable, in addition to mechanisms for redressal, safeguarding rights, and ensuring performance of corresponding duties.

Grounding ethical principles in human rights standards would also preclude the worrying prospect of "ethics washing" and rubber stamping. Human rights law has a legacy of constitutional and judicial interpretation, legal oversight, and enforcement mechanisms that are subject to review.

We also observed that industry initiatives and corporate aims have come to dominate and set the agenda within many ethics initiatives. A human rights-based approach could re-balance the scales, as it would embrace the realities of carrying out business by prescribing a specific, detailed account of what rights and obligations are in play. It would also keep individuals and the rights owed to them as the central focus, and the point for calibration.

Perhaps as a result of the agenda-setting by industry, the ethics framework does not contemplate the duty of companies to act in certain ways. Instead, it tends to focus on responsibility for outcomes (who holds it, and what it entails), as opposed to addressing bigger questions around business models and the monopolistic power of tech companies. A human rights-based approach would correct for this by targeting every step of a company's business model, thus bringing into focus not just outputs from a particular business, but also processes and safeguards for rights holders and duty bearers. This could also be drawn from the UNGPs, alongside international and national law, if applicable.

Finally, we touched upon the the risk (common in multi-stakeholder initiatives) of developing recommendations that are internally contradictory and do not provide balancing mechanisms for conflicting principles. A rights-based approach would provide guidance on how to balance principles as it would be informed by the reasonable restrictions on particular rights, the contours within which the rights stand, and how these conflicts have been treated in the law. Finding parallels in law can be immensely helpful.

3.2 A human rights-based approach to FAT

In the section discussing FAT above, we identified a few key issues where this work could be further developed to incorporate the social, political, and legal context in which AI systems are deployed.

A human rights-based approach could strongly complement the technical specificity of the FAT approach by providing a robust legal framework that is grounded in the relationship between rights holders and duty bearers, and articulates standards for enforcement, accountability, meaningful redressal, and individual empowerment. Technical definitions of fairness, transparency and accountability tend to focus mostly on outcomes in a given situation, not on the process of reaching the outcome. A rights-based approach would consider due process and non-discrimination in ways that transcend just outputs. ARTICLE 19 believes that a human rights-respecting AI system is not simply one that does not produce rights-violating outputs, but that it is human rights-respecting through design, development and deployment.

Invoking a human rights-based approach would allow FAT proponents to make a shift from precise technical models to fair sociotechnical systems. Human rights-based approaches encourage an understanding of context, rights, and corresponding duties in a manner that necessitates taking into account social realities and institutions. The human rights framework also provides a baseline acknowledgement of the scope and application of each right.

In turn, this could allow the FAT field to more meaningfully address the tangible effects of AI systems on the people's lives and fundamental rights. Because human rights-based approaches are grounded in the relationship between rights holders and duty bearers, it refocuses these questions in terms of individual and collective rights, and the obligations owed to rights holders, including rights to seek redressal.

The FAT approach does not empower individuals, or provide particular agency in this regard. A human rights-based approach offers enforcement mechanisms, and prescribes institutional processes to work with in the event of rights violations, in addition to offering existing external mechanisms. In case of discrimination, for instance, the FAT approach may lay out tests for fairness and methods of accountability, and a rights-based framework augments these by also providing redressal mechanisms for people affected.

The FAT approach focuses on systems, and mechanisms for accountability around systems, but does not clearly articulate responsibility, harm, and expectation of fair treatment. A human rights-based approach is particularly well positioned to fill this gap as it is informed by international legal standards and clear articulation of roles and responsibilities.

A human rights-based approach could also help to carry forward questions in the FAT field about how to engage with and critically examine values. Amid efforts to promote value-sensitive design[72] of machine learning systems, and a growing awareness of the need to explicitly deal with values that are encoded in systems, human rights provide a universal set of values with legal grounding. ARTICLE 19 believes that invoking a human rights framework is especially important given these learnings, as human rights are the most universal set of values that we have, with shared language and decades of interpretation and implementation.

Finally, a human rights-based approach would provide guidance on how to balance rights and values, or how to navigate tradeoffs between say, fairness and accuracy. Under a human rights-based approach, these would be informed by reasonable restrictions on particular rights, and how similar scenarios have been treated in case law.

3.3 How might FAT and ethics initiatives help improve a human rights-based approach?

Advocacy for human rights protections should take into account technical considerations: While human rights have grounding in law, and also universally understood language and meaning, advocates for human rights could strengthen their approach to AI systems by learning and taking into account technical necessities, limitations, and terminologies. This would not only enable precision across disciplines to emerge, it would also carry out important translation between technical and non-technical audiences.

A deeper understanding of technical considerations could help as well. There are some inevitable tradeoffs that must be made in the process of developing AI systems. For instance, the tradeoff between fairness and accuracy is a common dilemma. A human rights-based approach would be strengthened by understanding the considerations that go into such tradeoffs.

Finally, while human rights serve as a minimum requirement for AI systems to adhere to, a human rights-based approach could be strengthened by also taking into consideration aspirational goals that go beyond just this minimum requirement, as ethical principles do.

4. Recommendations

We call on industry to:

1. Affirm their commitment to the **UN Guiding Principles on business and human rights.**

2. Embed **international human rights standards**, particularly those of protecting freedom of expression and privacy, in the development and deployment of AI systems in a manner that lays out precise obligations and commitments for companies building these technologies.

3. Explicitly invoke a human rights-based approach to ethical codes of conduct; identify duty bearers and rights holders.

4. Publicly articulate what safeguards are in place to monitor and evaluate ongoing ethical efforts, and institute **accountability and redressal mechanisms.**

5. **Enhance transparency around ethical AI initiatives**, from the constitution of ethical boards (including their powers and functions), to how ethical principles are internally evaluated, to the impact of these initiatives on businesses.

6. Ensure that AI systems undergo rigorous **human rights impact assessments**, and create feedback and continuous auditing mechanisms for the same.

7. As a continuation of the above, **meaningfully engage with civil society and academia** at each stage of the process, to cultivate constructive criticism as part of internal deliberation processes.

We call on states to:

1. Reaffirm their **commitment to international human rights standards**, particularly those protecting freedom of expression and privacy, and let this commitment guide their specific national AI strategies.

2. Ensure that AI systems in the public sector undergo adequate **human rights impact assessments**, due diligence, and continuous auditing. These are not systems that can simply be rolled out. They should instead be tailored to the exact context and use for which they are intended.

3. Root the design, development, and deployment of AI systems in **constitutional guarantees and human rights standards.**

4. Hold AI systems to **accountability, responsibility, and constitutional standards without dilution or exception**.

5. Ensure that national and international efforts around AI are equally informed by **human rights concerns, constitutional standards, and the public interest** as they are by industry concerns.

6. Establish necessary infrastructure and resources to provide remedy for human rights violations caused by AI systems.

We call on civil society to:

1. Work towards greater **transparency and accountability** of mechanisms (such impact assessments, the Universal Periodic Review, and rights-based mechanisms at national levels) and institutions (like the Human Rights Council, General Assembly, Special Rapporteurs, national and regional institutions that focus on compliance with and enforcement of human rights) that make up a rights-based approach.

2. Advocate for the promotion and protection of human rights in the context of AI systems in a manner that is also informed by **technical considerations and limitations.**

3. Actively engage with ethical initiatives to integrate strong human rights language and protections into their substantive and procedural functioning.

Endnotes

1 In a previous report co-written with Privacy International, we attempted to capture this new technology in the following definition: "The term 'AI' is used to refer to a diverse range of applications and techniques, at different levels of complexity, autonomy and abstraction. This broad usage encompasses machine learning (which makes inferences, predictions and decisions about individuals), domain-specific AI algorithms, fully autonomous and connected objects and even the futuristic idea of an AI 'singularity'." We also outlined a number of key concepts related to AI systems. ARTICLE 19, and Privacy International. 'Privacy and Freedom of Expression In the Age of Artificial Intelligence', 2018. https://www.article19.org/wp-content/uploads/2018/04/Privacy-and-Freedom-of-Expression-In-the-Age-of-Artificial-Intelligence-1.pdf.

2 Vincent, James. 'Artificial Intelligence Is Going to Supercharge Surveillance'. The Verge, January 2018. https://www.theverge.com/2018/1/23/16907238/artificial-intelligence-surveillance-cameras-security.

3 Coughlan, Sean. 'Google "to End Pentagon AI Project"', June 2018, sec. Business. https://www.bbc.com/news/business-44341490.

4 Kania, Elsa B. 'China's AI Giants Can't Say No to the Party'. Foreign Policy (blog), August 2018. https://foreignpolicy.com/2018/08/02/chinas-ai-giants-cant-say-no-to-the-party/.

5 Barocas, Solon, and Andrew D. Selbst. 'Big Data's Disparate Impact'. California Law Review 671 (2016). https://papers.ssrn.com/sol3/Papers.cfm?abstract_id=2477899.

6 Veale Michael, Binns Reuben, and Edwards Lilian. 'Algorithms That Remember: Model Inversion Attacks and Data Protection Law'. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 376, no. 2133 (November 2018). https://doi.org/10.1098/rsta.2018.0083.

7 Cowls, Josh, and Luciano Floridi. 'Prolegomena to a White Paper on an Ethical Framework for a Good AI Society', (June 2018). https://papers.ssrn.com/abstract=3198732.

8 Floridi, et al. 'An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations'. In Minds and Machines (December 2018). https://www.researchgate.net/publication/328699738_An_Ethical_Framework_for_a_Good_AI_Society_Opportunities_Risks_Principles_and_Recommendations.

9 Dutton, Tim. 'An Overview of National AI Strategies'. Politics + AI (blog), June 2018. https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd.

10 'Ethics guidelines for Trustworthy AI'. Accessed 15 April 2019. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.

11 'Asilomar AI Principles'. Future of Life Institute. Accessed 11 April 2019. https://futureoflife.org/ai-principles/.

12 ARTICLE 19. 'Submission of Evidence to the House of Lords Select Committee on Artificial Intelligence', September 2017. https://www.article19.org/wp-content/uploads/2017/10/ARTICLE-19-Evidence-to-the-House-of-Lords-Select-Committee-AI-1.pdf. Also see 'High-Level Expert Group on Artificial Intelligence | Digital Single Market'. Accessed 11 April 2019. https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence.

13 'Partners - The Partnership on AI'. Accessed 11 April 2019. https://www.partnershiponai.org/partners/.

14 ARTICLE 19. 'Google: New Guiding Principles on AI show progress but still fall short on human rights protections', June 2018. https://www.article19.org/resources/google-new-guiding-principles-on-ai-show-progress-but-still-fall-short-on-human-rights-protections/.

15 Google. 'Our Principles'| Google AI. Accessed 11 April 2019. https://ai.google/principles/.

16 McNamara, Andrew, Justin Smith and Emerson Murphy Hill. 'Does ACM's Code of Ethics Change Ethical Decision Making in Software Development?. Accessed 15 April 2019. https://people.engr.ncsu.edu/ermurph3/papers/fse18nier.pdf.

17 Boyle, Alan. 'AI Expert Says Microsoft Is Cutting off Some Sales Due to Ethics Concerns – GeekWire'. Geekwire, April 2018. https://www.geekwire.com/2018/microsoft-cutting-off-sales-ai-ethics-top-researcher-eric-horvitz-says/.

18 Madrigal, Alexis C. 'What Facebook Did to American Democracy'. The Atlantic, October 2017. https://www.theatlantic.com/technology/archive/2017/10/what-facebook-did/542502/. Also see: Angwin, Julia, Madeleine Varner and Ariana Tobin. ' Facebook Enabled Advertisers to Reach 'Jew Haters''. ProPublica, September 2017. https://www.propublica.org/article/facebook-enabled-advertisers-to-reach-jew-haters.

19 'Facebook and the Technical University of Munich Announce New Independent TUM Institute for Ethics in Artificial Intelligence | Facebook Newsroom'. Accessed 11 April 2019. https://newsroom.fb.com/news/2019/01/tum-institute-for-ethics-in-ai/.

20 'About - The Partnership on AI'. Accessed 11 April 2019. https://www.partnershiponai.org/about/.

21 'Ro Khanna and Brenda l. Lawrence Introduce Resolution Calling for the Ethical Development of Artificial Intelligence'. Congressman Ro Khanna, February 2019. https://khanna.house.gov/media/press-releases/release-ro-khanna-and-brenda-l-lawrence-introduce-resolution-calling-ethical.

22 Ali, Muhammad, Piotr Sapiezynski, Miranda Bogen, Aleksandra Korolova, Alan Mislove, and Aaron Rieke. 'Discrimination through Optimization: How Facebook's Ad Delivery Can Lead to Skewed Outcomes'. ArXiv:1904.02095 [Cs], (April 2019. http://arxiv.org/abs/1904.02095.

23 'HUD versus Facebook Charge', August 2018. https://www.hud.gov/sites/dfiles/Main/documents/HUD_v_Facebook.pdf. Please note, Facebook recently settled these challenges. Please see: Levy, Pema. 'Facebook Settles Civil Rights Lawsuits Over Ad Discrimination'. Mother Jones, March 2019. https://www.motherjones.com/politics/2019/03/facebook-settles-civil-rights-lawsuits-over-ad-discrimination/.

24 Gallagher, Ryan. 'Google Plans to Launch Censored Search Engine in China, Leaked Documents Reveal'. The Intercept (blog), August 2018. https://theintercept.com/2018/08/01/google-china-search-engine-censorship/.

25 Cath, Corinne. 'Who Is Driving the AI Agenda and What Do They Stand to Gain?' NS Tech (blog), July 2018. https://tech.newstatesman.com/guest-opinion/regulating-artificial-intelligence-ai.

26 'High-Level Expert Group on Artificial Intelligence | Digital Single Market'. Accessed 11 April 2019. https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence.

27 Dialogue Seminar on Artificial Intelligence: Ethical Concerns - Comments by T. Metzinger, 2019. http://www.europarl.europa.eu/streaming.

28 'Members AI Task Force (AITF)'. Accessed 11 April 2019. https://www.aitf.org.in/members.

29 Williams, Oscar. 'DeepMind's Demis Hassabis Is Set to Advise the Government's New Office for AI'. NS Tech (blog), June 2018. https://tech.newstatesman.com/business/demis-hassabis-office-ai-adviser.

30 Wagner, Ben. 'Ethics as an Escape from Regulation: From ethics-washing to ethics-shopping?' . M. Hildebrandt (Ed.), Being Profiling. Cogitas ergo sum. Amsterdam University Press (2018). https://www.privacylab.at/wp-content/uploads/2018/07/Ben_Wagner_Ethics-as-an-Escape-from-Regulation_2018_BW9.pdf.

31 Nemitz, Paul. 'Constitutional Democracy and technology in the age of artificial intelligence'. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 376, no. 2133 (November 2018). https://doi.org/10.1098/rsta.2018.0089.

32 Kharpal, Arjun. 'Intel CEO Brian Krzanich: It's Too Early to Regulate AI', November 2017. https://www.cnbc.com/2017/11/07/ai-infancy-and-too-early-to-regulate-intel-ceo-brian-krzanich-says.html.

33 Smith, Reiss. 'Google appoints transphobic conservation to AI ethics board'. Pink News, March 2019. https://www.pinknews.co.uk/2019/03/27/google-ai-artificial-intelligence-ethics-board-kay-coles-james/.

34 Piper, Kelsey. 'Google cancels ethics board in response to outcry'. Vox, April 2019. https://www.vox.com/future-perfect/2019/4/4/18295933/google-cancels-ai-ethics-board.

35 Weissmann, Jordan. 'Amazon Created a Hiring Tool Using AI. It Immediately Started Discriminating Against Women.' Slate Magazine, October 2018. https://slate.com/business/2018/10/amazon-artificial-intelligence-hiring-discrimination-women.html.

36 Angwin, Julia et. al. 'Machine Bias'. ProPublica, May 2016. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

37 ARTICLE 19, and Privacy International. 'Privacy and Freedom of Expression In the Age of Artificial Intelligence', 2018. https://www.article19.org/wp-content/uploads/2018/04/Privacy-and-Freedom-of-Expression-In-the-Age-of-Artificial-Intelligence-1.pdf.

38 Narayanan, Aravind. 'FAT* Tutorial: 21 Fairness Definitions and Their Politics'. Google Docs, February 2018. https://docs.google.com/document/d/1bnQKzFAzCTcBcNvW5tsPuSDje8WWWY-SSF4wQm6TLvQ/edit?usp=sharing&usp=embed_facebook.

39 See, for instance, Kleinberg, Jon, Sendhil Mullainathan, and Manish Raghavan. 'Inherent Trade-Offs in the Fair Determination of Risk Scores'. ArXiv:1609.05807 [Cs, Stat], (September 2016). http://arxiv.org/abs/1609.05807.

40 Marda, Vidushi. 'Machine Learning and Transparency: A Scoping Exercise', (November 2017. https://papers.ssrn.com/abstract=3236837.

41 Diakopoulos, Nick and Michael Koliska. 'Algorithmic Transparency in the News Media'. In Digital Journalism (2016), DOI: 10.1080/21670811.2016.1208053.

42 Kroll, Joshua A, et al. 'Accountable Algorithms'. In University of Pennsylvania Law Review, Vol. 165, 2017 Forthcoming. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2765268.

43 ACM Conference on Fairness, Accountability, and Transparency (FAT*). Accessed 15 April, 2019. https://fatconference.org/.

44 Buolamwini, Joy, and Timnit Gebru. 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification'. In Conference on Fairness, Accountability and Transparency, 2018, 77–91. http://proceedings.mlr.press/v81/buolamwini18a.html.

45 Metcalf, Jacob. Engineering for Fairness: How a Firm Conceptual Distinction between Unfairness and Bias Makes it Easier to Address Un/Fairness. Conference on Fairness, Accountability, and. Transparency, New York,

Forthcoming Proceedings of Machine Learning Research (March 2018). http://ethicalresolve.com/content/uploads/2019/01/fat2019tutorials-paper26.pdf.

46 Babaei, Mahmoudreza et al, 'Analyzing Biases in Perception of Truth in News Stories and Their Implications for Fact Checking'. In FAT* '19 Proceedings of the Conference on Fairness, Accountability, and Transparency Pages 139-139 (2019). https://dl.acm.org/citation.cfm?id=3287581.

47 Lomas, Natasha. 'Accenture Wants to Beat Unfair AI with a Professional Toolkit'. TechCrunch, 2018. http://social.techcrunch.com/2018/06/09/accenture-wants-to-beat-unfair-ai-with-a-professional-toolkit/.

48 Cramer, Henriette, et al. 'Challenges of incorporating algorithmic 'fairness' into practice'. In Conference on Fairness Accountability, and Transparency Tutorial (February 2019). https://drive.google.com/file/d/1rUQkVS0NzSH3IEqZDsczSxBbhYHbjamN/view.

49 Marda, Vidushi. 'Artificial Intelligence Policy in India: A Framework for Engaging the Limits of Data-Driven Decision-Making'. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 376, no. 2133 (November 2018). https://doi.org/10.1098/rsta.2018.0087.

50 Binns, Reuben. 'Fairness in Machine Learning: Lessons from Political Philosophy'. Cnference on Fairness, Accountability, and. Transparency, New York, Forthcoming Proceedings of Machine Learning Research Vol. 81 (December 2017): 1–11. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3086546.

51 Miller, Tim. 'Explanation in Artificial Intelligence: Insights from the Social Sciences'. ArXiv:1706.07269 [Cs], (June 2017. http://arxiv.org/abs/1706.07269.

52 Hutchinson, Ben, and Margaret Mitchell. '50 Years of Test (Un) Fairness: Lessons for Machine Learning'. Proceedings of the Conference on Fairness, Accountability, and Transparency - FAT* '19, (2019, 49–58. https://doi.org/10.1145/3287560.3287600.

53 Buolamwini, Joy, and Timnit Gebru. 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification'. In Conference on Fairness, Accountability and Transparency, 2018, 77–91. http://proceedings.mlr.press/v81/buolamwini18a.html.

54 'Mayor de Blasio Announces First-In-Nation Task Force To Examine Automated Decision Systems Used By The City'. The Official Website of the City of New York, May 2018. https://www1.nyc.gov/office-of-the-mayor/news/251-18/mayor-de-blasio-first-in-nation-task-force-examine-automated-decision-systems-used-by.

55 Stoyanovich, Julia, and Solon Barocas. Testimony of Julia Stoyanovich and Solon Barocas before New York City Council Committee on Technology,

regarding Update on Local Law 49 of 2018 in Relation to Automated Decision Systems (ADS) Used by Agencies, § New York City Council Committee on Technology (/2019). https://dataresponsibly.github.io/documents/StoyanovichBarocas_April4,2019testimony.pdf.

56 Regulation (EU) 2016/679 General Data Protection Regulation (GDPR), https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A32016R0679.

57 Regulation (EU) 2016/679, Article 13(2)(f), 14(2)(g), 15(1)(h).

58 Regulation (EU) 2016/679, Article 22(1).

59 Regulation (EU) 2016/679, Article 22(1) - 22(4).

60 Wachter, Sandra, Brent Mittelstadt and Luciano Floridi. 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' In International Data Privacy Law (December 2017). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2903469. Also see Edward, Lilian, and Michael Veale. 'Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For'. In 16 Duke Law & Technology Review 18 (2017). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2972855. Also see Selbst, Andrew D. and Julia Powles. 'Meaningful Information and the Right to Explanation'. In International Data Privacy Law, vol. 7(4), 233-242 (2017). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3039125.

61 Algorithmic Accountability Act, 2019. Accessed April 15, 2019. https://www.wyden.senate.gov/imo/media/doc/Algorithmic%20Accountability%20Act%20of%202019%20Bill%20Text.pdf.

62 ARTICLE 19, and Privacy International. 'Privacy and Freedom of Expression In the Age of Artificial Intelligence', 2018. https://www.article19.org/wp-content/uploads/2018/04/Privacy-and-Freedom-of-Expression-In-the-Age-of-Artificial-Intelligence-1.pdf.

63 Latonero, Mark. 'Governing Artificial Intelligence - Upholding Human Rights and Dignity'. Data & Society, October 2018. https://datasociety.net/output/governing-artificial-intelligence/.

64 'The Toronto Declaration: Protecting the Rights to Equality and Non-Discrimination in Machine Learning Systems - Access Now'. Accessed 11 April 2019. https://www.accessnow.org/the-toronto-declaration-protecting-the-rights-to-equality-and-non-discrimination-in-machine-learning-systems/.

65 'Pretrial Risk Assessments'. The Leadership Conference Education Fund. Accessed 11 April 2019. https://civilrights.org/edfund/pretrial-risk-assessments/.

66 Kaye, David. 'OHCHR | Report of the Special Rapporteur to the General Assembly on AI and Its Impact on Freedom of Opinion and Expression' (United Nations Office of the High

Commissioner of Human Rights, 2018). https://www.ohchr.org/EN/Issues/FreedomOpinion/Pages/ReportGA73.aspx.

67 Australian Human Rights Commission. 'Protecting Human Rights in the Era of Artificial Intelligence', July 2018. https://www.humanrights.gov.au/news/stories/protecting-human-rights-era-artificial-intelligence.

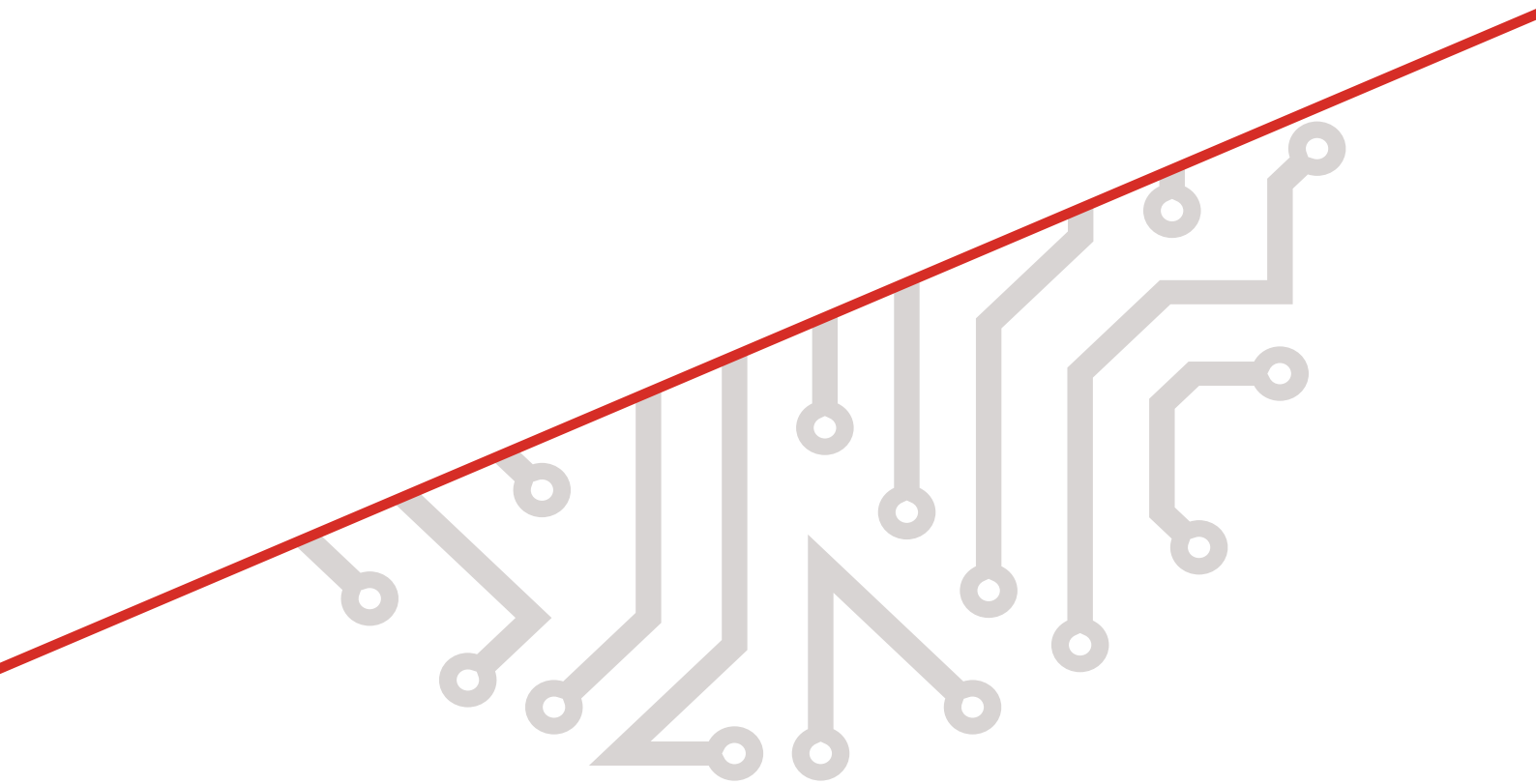
68 Simonite, Tom. 'Canada, France Plan Global Panel to Study the Effects of AI'. Wired, December 2018. https://www.wired.com/story/canada-france-plan-global-panel-study-ai/.

69 Allison-Hope, Dunstan and Hodge, Mark. 'Artificial Intelligence: A Rights-Based Blueprint for Business'. Business for Social Responsibility BSR, August 2018. https://www.bsr.org/reports/BSR-Artificial-Intelligence-A-Rights-Based-Blueprint-for-Business-Paper-01.pdf.

70 Veen, Christiaan van, and Corinne Cath. 'Artificial Intelligence: What's Human Rights Got To Do With It?' Data & Society: Points, May 2018. https://points.datasociety.net/artificial-intelligence-whats-human-rights-got-to-do-with-it-4622ec1566d5.

71 United Nations Human Rights Office of the Human Rights Commissioner. 'Guiding Principles on Business and Human Rights', 2011. https://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf.

72 Dobbe, Roel, and Morgan Ames. 'Up Next For FAT*: From Ethical Values To Ethical Practices'. Medium, February 2019. https://medium.com/@roeldobbe/up-next-for-fat-from-ethical-values-to-ethical-practices-ebbed9f6adee.



ARTICLE 19 Free Word Centre 60 Farringdon Road London EC1R 3GA
T +44 20 7324 2500 F +44 20 7490 0566
E info@article19.org W www.article19.org Tw @article19org facebook.com/article19org